A mutational analysis of the Bacillus subtilis competence helicase ComFA

A mutational analysis of the *Bacillus subtilis* competence helicase ComFA

Abstract

Genetic competence is a developmental process in bacteria that allows natural transformation. Competent Gram positive bacteria such as *Bacillus subtilis* carry a cytosolic helicase which is required for efficient transformation. In this work ComFA is confirmed as a DEAD-box helicase. I also describe a new accessory motif in ComFA that contributes to transformation independently of the helicase activity in ComFA. The newly discovered metal-binding motif consists of four cysteines which are required for transformation and zinc binding. While the zinc finger is required for full function, it is not required for DNA binding. As DEAD-box family helicases are generally non-processive, it appears that at least part of the rapid DNA uptake process is mediated by a non-processive helicase. Active uptake using the ComFA helicase motor may be required to maintain the integrity of the incoming DNA to allow subsequent recombination.

To my parents and family who have always supported and believed in my dreams.

Acknowledgments

The work which follows in this dissertation is the result of a number of years of difficult work continuing an examination of the *Bacillus subtilis* competence protein ComFA. This is not an exhaustive list, and I apologize to anyone I may have missed. I am sincerely grateful to everyone who has helped me along the way to completing this work.

I would like to say thank you to Dr. Briana Burton, and the members of the Burton Lab both, past and present. You have all made the lab an interesting place to work, and graduate school a unique experience. I would like to especially thank Mark Georgian, Anna Zwieniecka, and Dr. Rangapriya Sundararajan. Your help and support throughout my time in graduate school are greatly appreciated and were integral in making the completion of the work described here successful.

To the laboratory of Professor Richard Losick, you have all been great neighbors over the years. The continual exchange of scientific ideas, and experimental supplies has greatly enriched my work. Dr. Niels Bradshaw deserves particularly acknowledgement for all of his advice during this project with regard to methods for purifying ComFA. I am also extremely grateful to Dr. Matthew Cabeen. Dr. Cabeen's contribution was essential to the discovery of the zinc finger in ComFA. While still a graduate student at Yale, he provided me with a very generous gift of the purified FlAsH dye used for the microscopy analysis included in Appendix B. These experiments provided the initial clues that the metal-binding activity existed in ComFA. Professor Rachelle Gaudet and members of her laboratory were also very helpful along the way while trying to purify ComFA. I am especially grateful to Dr. Wilhelm Weihofen. He provided me with the constructs and the protocol needed to purify the PreScission Protease used to cleave the partially purified MBP-ComFA, which is part of the work presented in Chapter 3. I thank all of the members of my Dissertation Advisory Committee during the entire time of my graduate studies. I thank Dr. Nicole Francis, Professor Daniel Kahne, and Professor Guido Guidotti for their guidance and service as members of my Dissertation Advisory Committee. I also thank Professor Richard Losick

and Professor Rachelle Gaudet for their service and guidance, as well as their chairmanship of my Dissertation Advisory Committee.

The members of the Department of Molecular and Cellular Biology Graduate Programs Office, both past and present, have provided a great deal of assistance during my time as a graduate student. They have been instrumental in facilitating the scheduling of regular meetings with my dissertation advisory committee and answering my questions about program requirements. They have also assisted greatly with finding resources in the department and around Harvard that have augmented my graduate school experience.

Professor Cassandra Extavour was a very helpful mentor to me during graduate school. I am very thankful to her for her investment in my work and my development as a scientist. She provided important advice and various point throughout my studies which greatly contributed to my scientific, professional, and personal development.

Thank you the members of my family who have always been there to support me throughout this process. Even though you have been far away, having you with me during the completion of my dissertation and through life is something I am eternally grateful for.

I have been blessed to find a great community of and develop a great group of friends during my time in graduate school. We have been brought together through a myriad of events and activities, and having your support means a lot to me. Many of you were around at some of the more difficult points in completing this work, and your help and support really enabled me to push through and continue making progress all along the way.

The Microbial Sciences Initiative Graduate Student Consortium has provided a great and vibrant community of scientists with a passion for microbiology. This community was a great source of important scientific discussions and diverse experiences which helped morph how I think about my work and how I approached my experiments and life beyond the laboratory. Participating in student government was also an important part of my graduate experience, and I thank everyone I worked with on the Harvard Graduate Council and the Graduate School of Arts and Sciences Graduate Student Council, for all your hard work in service to your fellow students, and providing me with a great experience which has greatly impacted how I think about my science, and think about my life.

I am extremely grateful to the other Howard Hughes Medical Institute Gilliam Fellows. You are an amazing group that I am proud to be associated with. The drive you all show in the pursuit of your goals, and mutual support you provide to other fellows has profoundly

impacted my life. I thank all of you for your suggestions, advice, lively scientific discussion, and for many of you, your friendship as well.

Finally, I would like to acknowledge some external sources of support. I received financial support from the Harvard University Graduate Science Fellowship, and the Howard Hughes Medical Institute Gilliam Fellowship. The HHMI Gilliam Fellowship Program also provided a great deal of professional and personal support. I am very grateful for all the efforts of the members of the HHMI Science Education group, especially, Dr. William Galey, Dr. Maryrose Franko, Andrew Quon, Christie Schultz, Leni Bautista, Makeda Richardson, and Megan Lassig. The work presented here was also supported by a grant from the Rita Allen Foundation obtained by Dr. Briana Burton. I would also like to acknowledge the thousands of individuals who have coded for the LaTeX project for free. It is due to their efforts that we can generate professionally typeset PDFs now. Typesetting was essential to the aestetics and quality of the work presented here.

Contents

Copyright ii

Abstract iii

Acknowledgements v

Contents viii

List of Figures xiii

List of Tables xv

1 Chapter 1: Introduction to genetic competence 1

1.1 Molecular transport . . . 1
1.2 Macromolecular transport . . . 2
1.3 Horizontal gene transfer . . . 5
1.4 Competence and transformation . . . 8
1.5 Competence in *Bacillus subtilis* . . . 9
1.5.1 Development of competence . . . 9
1.5.2 Transformation in *B. subtilis* . . . 11
1.5.2.1 DNA binding . . . 11
1.5.2.2 Fragmentation . . . 12
1.5.2.3 Uptake and internalization . . . 12
1.5.2.4 Resolution . . . 13
1.5.3 The competence machinery . . . 13

2 Mutational analyisis of the ComFA canonical DEAD-box helicase motifs 16
2.1 Abstract . . . 16
2.2 Introduction . . . 17
2.2.1 The helicase superfamilies . . . 18
2.2.2 DEAD-box helicases . . . 19
2.2.3 ComFA . . . 20
2.3 Results . . . 24
2.3.1 ComFA is a DEAD-box helicase . . . 24
2.4 Conclusions . . . 25
2.5 Materials and methods . . . 30
2.5.1 Strains and growth conditions . . . 30
2.5.2 Plasmid construction . . . 30
2.5.3 Construction of *B. subtilis* strains . . . 35
2.5.3.1 *comF* expression strains . . . 35
2.5.3.2 Genomic DNA preparation and PCR amplification . . . 36
2.5.3.3 Colony PCR . . . 37
2.5.4 Transformation efficiency analysis . . . 37
2.5.4.1 *comF* expression strains . . . 37
2.5.5 Statistical analysis . . . 38

3 Chapter 3: Characterization of a zinc finger in ComFA 44
3.1 Abstract . . . 44
3.2 Introduction . . . 45
3.2.1 Metal-binding and biochemistry . . . 45
3.2.2 Zinc and transition metals . . . 45
3.2.3 The putative zinc finger in ComFA . . . 47
3.3 Results . . . 48
3.3.1 Discovery of possible metal-binding activity . . . 48
3.3.2 Bioinformatic analysis . . . 48
3.3.3 Zinc finger is important for ComFA function . . . 51
3.3.4 The zinc-binding activity has an independent contribution to ComFA function from the DEAD-box helicase motifs . . . 51

3.3.5 Purification of ComFA . . . 53
3.3.5.1 Purification of H_6-ComFA . . . 55
3.3.5.2 Purification of MBP-ComFA . . . 55
3.3.6 ComFA binds zinc . . . 62
3.3.7 Tetracysteine motif required for zinc-binding activity . . . 64
3.3.8 On-going work . . . 67
3.4 Conclusions . . . 67
3.5 Materials and methods . . . 72
3.5.1 Strains and growth conditions . . . 72
3.5.2 Plasmid construction . . . 72
3.5.3 Construction of *B. subtilis* strains . . . 79
3.5.3.1 Haploid *yvbJ* expression strains . . . 79
3.5.3.2 *comF* expression strains . . . 79
3.5.4 Transformation efficiency analysis . . . 80
3.5.4.1 *yvbJ* expression strains . . . 80
3.5.4.2 *comF* expression strains . . . 80
3.5.4.3 Genomic DNA preparation and PCR amplification . . . 80
3.5.4.4 Colony PCR . . . 80
3.5.4.5 Sequencing to verify mutants . . . 80
3.5.5 Statistical Analysis . . . 80
3.5.6 Purification of PreScission Protease . . . 81
3.5.7 Expression of MBP-ComFA . . . 82
3.5.7.1 Enterokinase construct . . . 82
3.5.7.2 Expression for zinc-IMAC analysis . . . 82
3.5.7.3 Expression for PAR analysis . . . 83
3.5.8 Purification of ComFA . . . 83
3.5.8.1 Enterokinase construct . . . 83
3.5.8.2 For zinc-IMAC analysis . . . 84
3.5.8.3 For PAR analysis . . . 84
3.5.9 Enterokinase proteolysis . . . 85
3.5.10 PreScission Protease proteolysis . . . 85
3.5.11 Zinc-IMAC analysis . . . 86

3.5.12 EMSA analysis . . . 86
3.5.13 PAR colormetric analysis . . . 86

4 Discussion **92**
4.1 Broader significance . . . 92
4.2 The importance of competence . . . 92
4.2.1 DNA as food . . . 93
4.2.2 DNA uptake for repair . . . 94
4.2.3 DNA uptake for genetic exchange . . . 95
4.3 ComFA and transformation . . . 99
4.4 ComFA is a DEAD-box helicase with a metal-binding motif . . . 101
4.5 Future directions and next steps . . . 104
4.5.1 Further analysis of ComFA . . . 104
4.5.1.1 *In vivo* experiments . . . 104
4.5.1.2 *In vitro* experiments . . . 105
4.6 Other members of the *comF* operon . . . 106
4.6.0.3 Analysis of *comFC* . . . 106
4.7 Final Thoughts . . . 107
4.7.1 Applications for natural transformation . . . 107

A Appendix A: Analysis of expression of *comFA* and mutants at ectopic locus *yvbJ* **108**
A.1 Canonical DEAD-box mutations are not dominant . . . 108
A.2 Expression from *yvbJ* creates a competence defect . . . 111
A.3 Canonical DEAD-box motifs are required for ComFA function . . . 114
A.4 Analysis of the ABC signature-like motif . . . 114
A.5 Conclusion . . . 120
A.6 Materials and methods . . . 124
A.6.1 Strains and growth conditions . . . 124
A.6.2 Plasmid construction . . . 124
A.6.3 *B. subtilis* strain construction . . . 130
A.6.3.1 Meridiploid strains . . . 130
A.6.3.2 Haploid *yvbJ* expression strains . . . 131

A.6.4 Transformation efficiency 131

B Appendix B: Discovery of zinc finger motif 138

B.1 Introduction 138

B.1.1 Localization of competence components 138

B.1.2 ComFA fluorescent tagging 139

B.1.2.1 Carboxy–terminal GFP is cleaved from ComFA 139

B.2 *comF* gene product binds FlAsH 142

B.3 Conclusion 145

B.4 Materials and methods 146

B.4.1 Strains and growth conditions 146

B.4.2 Plasmid construction 146

B.4.3 ComFA-GFP expression time course 148

B.4.4 Fluorescence microscopy 148

B.4.4.1 FlAsH 148

B.4.4.2 Image analysis 149

References 152

List of Figures

1.1 Transport mechanisms 3
1.2 Conjugation, transduction, and transformation 7
1.3 Development of Competence in *B. subtilis* 10
1.4 Model of Transformation in *B. subtilis* 15

2.1 Schematic of ComFA 22
2.2 ComFA resembles a DEAD-box protein 23
2.3 DEAD-box motifs are required for ComFA function 26

3.1 ComFA contains a conserved tetracysteine motif 50
3.2 Cysteines are required for efficient transformation and zinc binding 52
3.3 The C4-zinc finger requirement is independent of the DEAD-box motif requirement 54
3.4 H_6-ComFA is insoluble 56
3.5 MBP-ComFA Enterokinase Cleavage 57
3.6 Purification of PreScission Protease 59
3.7 Zinc improves the solubility of MBP-ComFA 60
3.8 DNA binding to purified ComFA 61
3.9 MBP-ComFA complements *comF* phenotype 63
3.10 Purification of MBP-ComFA for Zn-IMAC 65
3.11 Tetracysteine motif is required for binding Zn-IMAC column 66
3.12 MBP-ComFA and MBP-ComFA^{4CS} purifications 68
3.13 MBP-ComFA and MBP-ComFA^{4CS} DNA-binding 69

4.1 Clonal interference 98

4.2 Model of ComFA mutant function during transformation 103

A.1 DEAD-box mutations are not dominant 110
A.2 *yvbJ::comFA* and *yvbJ::comFABC* expression do not achieve wild type transformation efficiency 113
A.3 Canonical DEAD-box motifs are required for ComFA function when expressed from ectopic locus 115
A.4 Schematic of ComFA motif compared with DEAD Family and ABC Family proteins 117
A.5 The ABC signature-like motif is not required for efficient transformation . . 119

B.1 Visualizing ComFA localization by fluorescence microscopy 140
B.2 Carboxy-terminal GFP is cleaved from ComFA 141
B.3 GFP-ComFA does not complement 142
B.4 FlAsH cross-reacts with *comF* gene product 144

List of Tables

2.1 Strains used in Chapter 2 39
2.2 Plasmids used in Chapter 2 40
2.3 Oligonucleotides used in Chapter 2 41

3.1 Strains used in Chapter 3 87
3.2 DNA & Plasmids used in Chapter 3 88
3.3 Oligonucleotides used in Chapter 3 90

A.1 Strains used in Appendix A 133
A.2 DNA & Plasmids used in Appendix A 134
A.3 Oligonucleotides used in Appendix A 136

B.1 ComFA and TC tag sequences 145
B.2 Strains used in Appendix B 149
B.3 DNA & Plasmids used in Appendix B 150
B.4 Oligonucleotides used in Appendix B 151

1

Chapter 1: Introduction to genetic competence

1.1 Molecular transport

Membranes are an integral part of cellular life. They function as semi-permeable barriers between the contents of the cell, and the external environment. Creating this spatial separation makes possible a number of biological processes. As such, cells have evolved a plethora of methods to control the flow of materials across cell membranes.

Transport mechanisms can be separated into two main classes, passive transport, and active transport, based upon the requirement for energy to transport a substrate across the membrane. Passive transport generally involves molecules moving down concentration gradients. Substances that have the ability to diffuse across the membrane do so freely. Molecules that cannot cross the membrane due to their charge or size undergo facilitated transport using protein channels embedded in the membrane (Figure 1.1A).

Active transport encompasses mechanisms that require energy to allow transport. Often these processes involve the import or export of substrates against their concentration gradients, or transport of very large substrates, such as macromolecules (Figure 1.1B & C). Macromolecular transport encompass the trafficking of biopolymers and other large secondary metabolites such as proteins, genetic material, and other membrane impermeable substrates.

1.2 Macromolecular transport

Both eukaryotes and bacteria have developed a number of ways of moving around large cargo. Generally, these fall into two main strategies: vesicular transport (endocytosis and exocytosis), in which the cargo is contained in membrane-bound compartments, and using specialized complexes which use channels and pumps to move macromolecules.

Vesicular transport is primarily a eukaryotic phenomenon (1). We often see this with the processing of proteins as they move from the endoplasmic reticulum (ER) to the Golgi apparatus, and then to the cell membrane, other cellular compartments, or are exported following fusion of the vesicle with the cell membrane (Figure 1.1C).

We see protein-based transport processes in both eukarya and bacteria. In eukarya the best characterized systems are for nucleic acid transport into and out of the nucleus, and polypeptide transport to the ER by the secretion machinery (2, 3). Bacteria have evolved a number of systems for export of proteins and nucleic acids. Scientists have identified at least

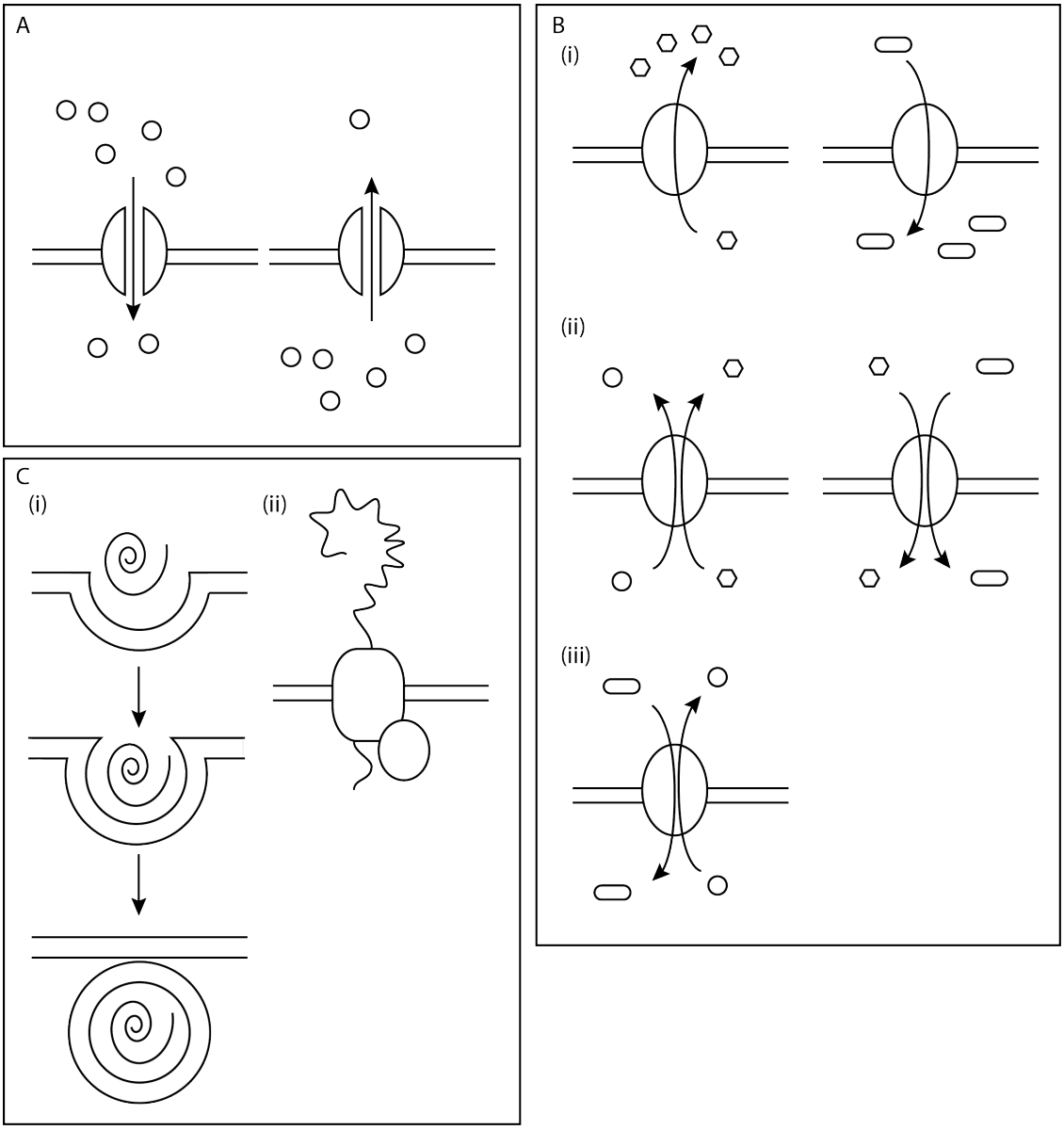

Figure 1.1: Transport mechanisms - (A)Passive Transport. A channel mediates the diffusion of a solute across the membrane, down a concentration gradient. (B)Active Transport (i) Membrane-bound pumps move substrates across the membrane. (ii) Symport. Two substrates are moved in the same direction across the membrane. Co-transport enables the energetics to transport the substrates. (iii) Antiport. Export of one substrate mediates the import of another. (C) Macromolecular Transport. (i) Vesicular transport. The incoming substrate is enveloped in a membrane as it is imported. (ii) Protein transport mediated by a secretion system. A channel, with an associated pump moves the macromolecule across the membrane.

seven types of secretion systems for protein export. Each of these systems are designed for handling substrates for specific functions and destinations outside of the cytoplasm (reviewed in (4)).

Nucliec acid transport, as with other macromolecule occurs both as import and export. The mechanisms of nucleic acid export are primarily governed by selfish elements such as conjugation elements and bacteriophage viral infections. Generally, the export mechanisms are more designed for the propagation of the conjugal elements or the phage, rather than for the benefit of the cell. However, there is at least one example of conjugation being used as a virulence method (5). There is evidence that nucleic acids are exported by some bacterial species by other mechanisms, however with the exception of export systems related to cell division and lysis, we know very little about how bacterial cells release DNA into the environment (6, 7, 8).

Bacteria have also developed methods for importing nucleic acids reviews in ((6, 9, 10)). These methods of genetic exchange allow for horizontal gene transfer, and are thought to be a source for genetic diversity in microorganisms. The evolutionary reasons for nucleic acid import is still a topic of great debate, and understanding how the transport systems function may shed some light on their value and utility (10, 11).

1.3 Horizontal gene transfer

Usually when we think about transmission of genetic material, passage of genes from parent to its progeny comes to mind. The phenomena which result in the inheritance of traits are referred to as vertical gene transfer (VGT). Coarse understanding of this process pre-dates the discovery of genes, and genetic theories (12). Overtime, our understanding of VGT has developed greatly and is the basis for the study of genetics.

There is another set of processes in which organisms can exchange genetic material in a non-hereditary manner, known as horizontal gene transfer (HGT). During HGT individuals in a community can exchange genetic material. The transfer may not be limited to members of the same species, depending on the mechanism of transfer. Genes acquired via HGT can be propagated vertically. There are three main natural mechanisms for HGT (Figure 1.2), and a number of methods have been developed to artificially transfer genetic material into some organisms (13, 14, 15, 16, 17) For the purposes of this discussion we will stick with the natural processes.

Natural processes that facilitate HGT are composed of conjugation, transduction, and transformation. Conjugation is sometimes referred to as bacterial sex, and involves direct transfer of mobile genetic elements (reviewed in (18)). Transduction refers to a transfer process which utilizes an intermediary. In bacteria, this is often a phage virion which captures part of the host genome (Figure 1.2Aiii & Biii) (19, 20). Transformation involves the

acquisition of free DNA from the environment. Topologically, conjugation and transduction are export processes meant to benefit selfish biological elements. Transformation, however, is topologically an import process in which a cell scavenges DNA from the environment (reviewed (21)).

Conjugation is an HGT process through which genetic material is transferred by a proteinaceous channel during cell-cell contact (Figure 1.2Ai & Bi). Often the genes encoding the export machinery are encoded on a plasmid, which in some cases integrates into the genome, and is propagated vertically. Under certain conditions the conjugation machinery can be induced which leads to replication of the plasmid, and transfer of the conjugative element to conjugation-deficient recipient cells (reviewed in (18)). The conjugation machinery in some cases can transfer other plasmids lacking the conjugation machinery also carried by the host cell (22). When a conjugative plasmid is transferred to a recipient cell, the recipient then becomes conjugation-capable. Due to the method of propagation, a conjugative element can be thought of as a paired-down lysogenic bacteriophage, which is able to propagate without lysing the host.

Transduction is a process through which genetic material is transferred from a donor cell to a recipient cell via an intermediary. In bacteria we see this most often with bacteriophage. Some phages make mistakes during the packing of new virions, and will package fragments of the host genome into some of the viral capsids, rather than a newly synthesized phage genome. The resulting virions will then transfer the region of the prior host's genome when

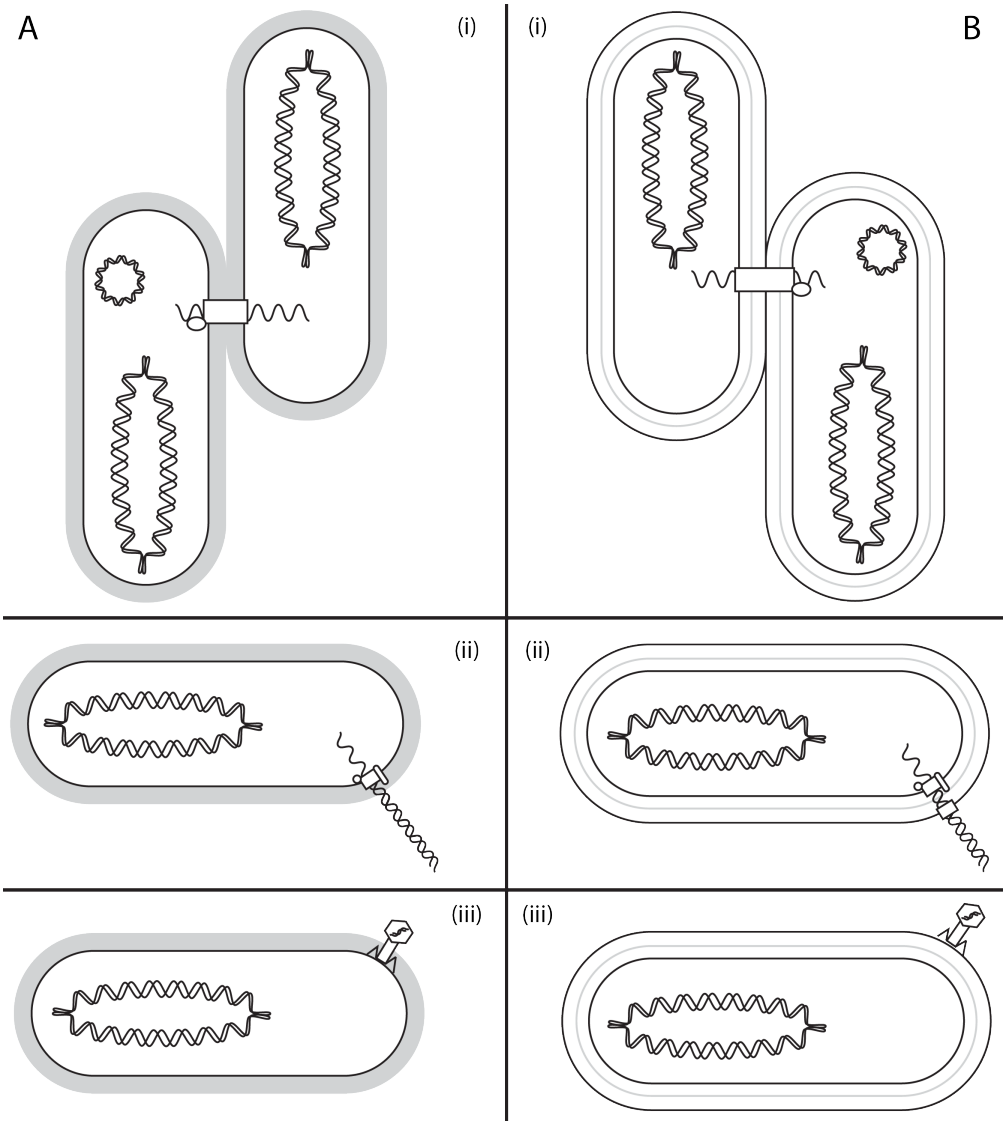

Figure 1.2: Conjugation, transduction, and transformation - (A) HGT mechanisms in Gram$^+$ bacteria (B) HGT Mechanisms in Gram$^-$ bacteria.(A and B)(i) Conjugation. DNA is transferred directly from the cytoplasm of one cell into another. (ii) Transformation. DNA is transported from the outside environment into the cell. (iii) Transduction. DNA from a bacteriophage is injected into a host cell. Gray region indicates peptidoglycan layers in each type of bacteria.

infecting a new host. If there are regions of homology between the region transferred and the new host genome, the region can be integrated into the new host genome and result in the loss or acquisition of genes (reviewd in (23)).

Transformation is the process by which cells take up DNA from their environment. Various species have mechanisms for mediating this process and utilizing the transported DNA molecules acquired by transformation. Transformation is of particular interest since it is an import process and is apparently driven by the cell, rather than by other elements.

1.4 Competence and transformation

Genetic competence (hereafter referred to as competence) is a process through which a cell develops the ability to take up DNA from its environment and become transformed. Our current understanding of natural competence is that it is a developmental state observed in both Gram positive (Gm^+) and Gram negative (Gm^-) bacteria, and involves the assembly of large protein complexes to allow passage of DNA across the cell wall and cell membranes (reviewed (21)). A number of elaborations on this theme have evolved as not all species use the same cues for development of competence. Due to the variations in regulation the timing of competence development varies greatly between species (24, 25) (reviewed in (25). Furthermore, the promiscuity or donor-range for the competence systems vary as well, with some species only allowing for intraspecies exchange, and others being receptive to DNA from any source (26, 27, 28, 29, 30, 31).

1.5 Competence in *Bacillus subtilis*

Competence has been well studied in *B. subtilis*, *Haemophilus influenza*, *Helicobacter pylori*, *Neisseria gonhorreae*, and *Streptococcus pneumoniae*. In the work discussed here *B. subtilis* was chosen as the studied system. We have a grasp of the proteins required for development of competence, as well as the genes required for DNA uptake, and transformation in all of the above species, however, competence and transformation have been most extensively studied in *B. subtilis* (reviewed in (10, 21, 32, 33, 34, 35)). *B. subtilis* is also an ideal system for studying transformation as it is non-pathogenic and has a wide DNA donor range (31).

1.5.1 Development of competence

Competence in *B. subtilis* develops as a post-exponential phase state. The master regulator *comK* is required for the expression of all the late competence (*com*) genes (36). ComK levels are controlled by a positive-feedback/proteolysis regulatory system. ComK is always expressed, but it is rapidly degraded by the MecA-ClpCP complex (37). ComS expression interferes with the MecA activity and allows for ComK levels to increase (37). This creates a bistable system in which competence arises stochastically in members of the population (38). The increase in ComK also allows expression of the late *com* genes (Figure 1.3) (36). In *B. subtilis* about 10 % of the cells can become competent in a stationary phase population (39, 40).

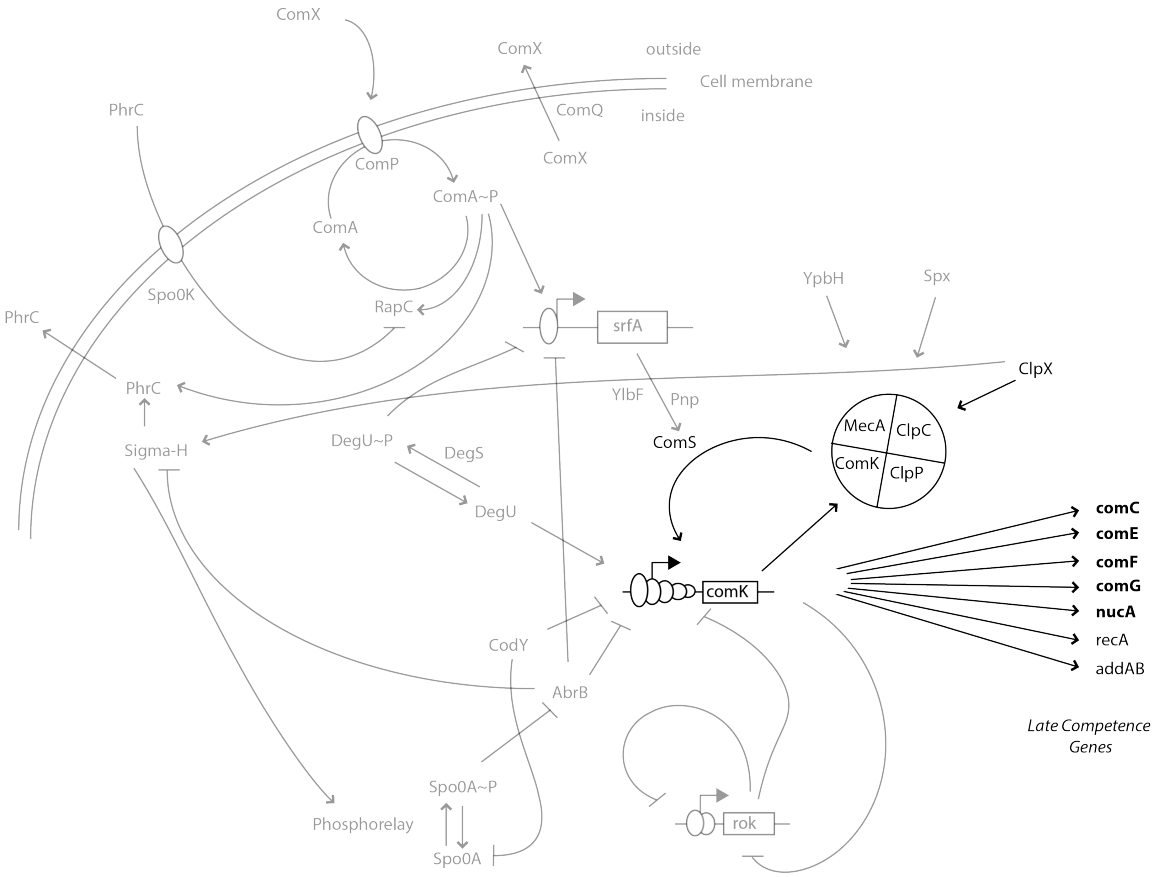

Figure 1.3: Development of Competence in *B. subtilis* - Diagram of the regulatory inputs that govern ComK and late competence gene expression. The late competence genes are expressed in cells to facilitate the transformation process. The components involved in the early regulation of *comK* have been marked in gray. Their involvement in the development of competence are reviewed in (33, 41). Adapted from(41).

1.5.2 Transformation in *B. subtilis*

A cell becomes competent when it has expressed the genes needed to undergo transformation. Transformation in *B. subtilis* occurs through a number of steps: DNA binding, fragmentation, uptake and internalization, and integration and resolution.

1.5.2.1 DNA binding

Binding is general defined as a non-covalent, wash-resistant attachment state for the transforming DNA (33). Work examining the kinetics of binding show that it occurs upon contact of the DNA substrate with the competent cell (42, 43). The DNA-binding activity occurs only in competent cells, and there appear to be a finite number of binding sites per cell, currently estimated to be about 50 sites per cell (42, 44). The quantity of DNA bound to a cells is proportional to the size of the polynucleotide bound, suggesting that DNA fragments are specifically bound to each site, rather than being bound as bulk masses of DNA (33). The attachment sites, and which cellular structures are responsible for initial binding are unknown, but must either transverse the cell wall, or be otherwise present on the outside of the cell to facilitate binding. Binding has also been shown to be DNA-specific, with little to no binding from double-stranded RNA or a number of synthetic polymers (28). This specificity will be important later when I discuss the purpose of competence.

1.5.2.2 Fragmentation

Fragmentation is a process through which double-strand breaks are introduced in the transforming DNA. The double-strand breaks are created by an extracellular endonuclease, NucA (45). The estimated sizes of the fragments observed are consistent with fragmentation along the cell length and width (33). That said, the sites of cleavage do not appear to be correlated with specific sequences or locations along the bound DNA (33).

1.5.2.3 Uptake and internalization

For a long time uptake in Gm^+ bacteria has been thought of as a single step, which was monitored by the conversion of the exogenous DNA to a DNase-resistant state. However, it appears that the DNase-resistant property may apply to two states of the incoming DNA (46, 47, 48). The first, being a movement which sequesters the DNA away from the cell surface, and a second which reflects the transport of the DNA across the cell membrane. This distinction of states has been observed in the delay between DNase resistance of the substrate and cyanide resistance of import during transformation (47), and the loss of protoplast associated DNA in *comGA* mutants (48). Uptake and internalization are thought to be driven by two factors, the proton motive force, which drives the collapse of the psuedopilus, and the cytosolic competence helicase ComFA (Figure 1.4) (43, 49).

1.5.2.4 Resolution

As the transforming strand is drawn into the cell it is bound by a number of single-strand binding proteins (50). The proteins presumably protect the transforming DNA from the nucleases present in the cytoplasm. The recombination (*rec*) machinery takes over at this stage and attempts to resolve the linear product of the transformation process into a circular element (31). This can happen in a couple ways. The linear fragment will be integrated into the genome if there is sufficient homology with a region of the genome, or can be recombined to create an extra-chromosomal element if there are sufficient regions of homology intramolecularly, or if multiple copies of the fragment are taken up, which is what we often see for plasmids (31, 51). As part of this recombination process the single-stranded fragment is converted back to being double-stranded by polymerization of the complementary strand.

1.5.3 The competence machinery

The DNA uptake machine is primarily composed of a number of late *com* genes. The current model for DNA uptake in *B. subtilis* com system begins when environmental DNA becomes bound to the surface of a competent cell. The ComG pseudo-pilins are type IV pilin-like proteins that form a tube that allows the incoming DNA to cross the cell wall and bind to ComEA (52, 53, 54). ComEA is a single-pass integral membrane protein containing an extracellular DNA-binding domain. ComEA has been demonstrated to have high DNA

binding affinity and is required for binding and transport even in the absence of a cell wall (54, 55, 56). Then, presumably, ComEA passes the incoming DNA across the cell membrane through ComEC via an undetermined mechanism. ComEC is a polytopic integral membrane protein that is thought to dimerize, forming an aqueous pore through the cell membrane, which allows the incoming DNA to cross the cell membrane (57). An unidentified nuclease degrades the non-transforming DNA strand and the degradation products are released into the environment (47). It is thought that meanwhile ComFA, a membrane-associated DEAD-box helicase pulls the transforming strand into the cell (49, 58, 59, 60). The *rec* machinery then binds the single-stranded transforming DNA strand to allow integration into the genome or formation of an extra-chromosomal DNA (31) (Figure 1.4).

The work discussed here examines a protein intimately involved in the import of the transforming strand during competence in *B. subtilis*, ComFA. ComFA appears to be the cytoplasmic motor of a very processive import machine, yet is predicted to be non-processive based upon its amino acid sequence. I examine the conserved helicase amino acid motifs predicted in the protein's primary sequence for their contribution to transformation, and also find an additional motif that appears to further contribute to activity.

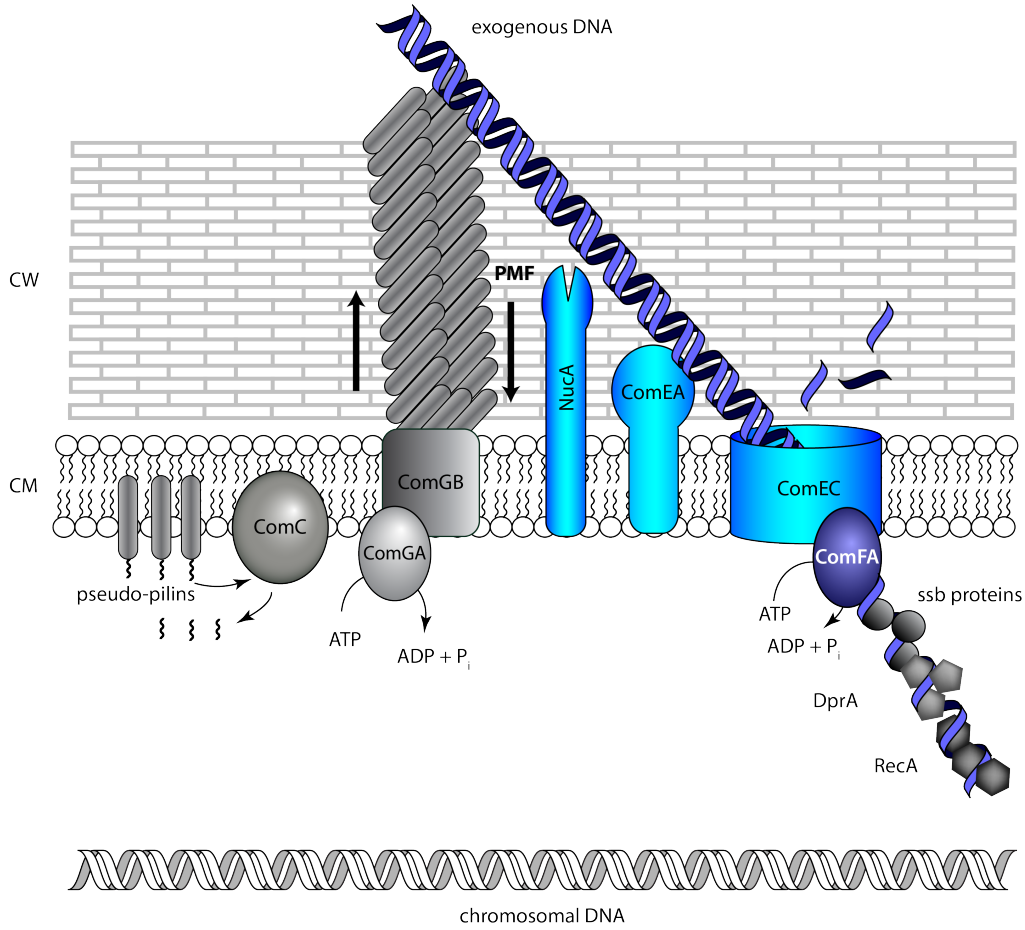

Figure 1.4: Model of Transformation in *B. subtilis* - DNA is brought into the cell by an initial binding to the outside of the cell mediated by the ComG pseudopilus. The proton motive force (PMF) is believed to drive the collapse of the pseudopilus bringing the incoming DNA to the cell membrane. A membrane-bound endonuclease (NucA)nicks the incoming DNA to create a free-end. The incoming DNA is bound by ComEA, and is pulled into the cell through ComEC by ComFA. During this process one DNA strand is degraded. The single-stranded transforming DNA is coated by single-stranded DNA binding proteins (ssb proteins) and the *rec* machinery (DprA and RecA). The coated transforming DNA is transported to the chromosome for homologous recombination. ComEA the primary, DNA receptor, ComEC the aqueous pore, and ComFA are labeled in shades of blue and purple. The incoming DNA is labeled in purple and black. The components which form the ComG pseudopilus, the single-stranded binding proteins, and the *rec* machinery are labeled in gray. The chromosomal DNA is labeled in white. CW: Cell wall, CM: Cell membrane. Adapted from (10, 21). NucA added based upon (45).

2

Mutational analyisis of the ComFA canonical DEAD-box helicase motifs

2.1 Abstract

DNA uptake during transformation in *B. subtilis* is understood to be a rapid and processive process. While it appears that some of the processivity behavior can be explained by the structural dynamics of the ComG pseudopilis and the proton motive force (43), defects observed in the DNA uptake process in ComFA mutants indicate that ComFA also has an important role [refs]. Bioinformatic analysis suggests that ComFA is a DEAD-box helicase (58). In the work presented in this chapter, I performed a mutational analysis of the conserved DEAD-box motifs in ComFA to confirm its designation as a DEAD-box helicase. I found that the conserved motifs are required for ComFA function. However, mutations in the conserved motifs create a 100–fold defect in tranformation efficiency which is less severe than the nearly 10 000-fold defect caused by the $comFA^{\Delta S1}$ in-frame deletion that was pre-

viously developed as a *comFA* null allele. ComFA is confirmed as a DEAD-box helicase, yet the helicase activity does not account for all of its contribution to efficient transformation.

2.2 Introduction

Genetic competence in *B. subtilis* is a developmental process in which a population of cells expresses genes that allow them to take up large stretches of DNA, up to 15 kb in size and integrate some portion of that DNA into their genome via homologous recombination (61). Superfamily II helicases and translocases are found in various DNA and RNA metabolic processes. ComFA, a member of the DEAD-box helicase family from sequence homology, is thought to be a major contributor to DNA uptake in *B. subtilis* (49, 58). As members of the DEAD-box helicase are not generally very processive, operating on substrate of 10—100 bp in length (reviewed in (62, 63)) it is interesting to understand how ComFA achieves this feat. I am interested in understanding how ComFA functions as a transport helicase, and how the mechanism of its function contributes to the DNA uptake process in *B. subtilis*. Genetic competence is observed in several other bacterial species, including human pathogens and the genes important in the process are conserved across the bacterial domain of life (6, 21, 64).

2.2.1 The helicase superfamilies

Nucleic acid helicases and translocases are central to life. They are required for DNA replication, RNA maintenance, sporulation, and horizontal gene transfer events in bacteria ((65), reviewed in (35)). They are also involved in a large number of regulatory processes in eukarya (62, 63, 66, 67, 68). These helicases are divided into six superfamilies (SF1—SF6) based on sequence homology, motifs, and structural features that define each superfamily. SF1 & 2 are related, and the member helicases primarily function as monomers and dimers. SF3—SF6 are made up of helicases that generally form rings and function as hexamers (reviewed in (63)). The quaternary structural differences between SF1 & SF2, and SF3—SF6 helicases translate into very different behaviors. For example, many of the SF1 & SF2 helicases are involved in low-processivity processes, in which they do not need to move along nucleic acid substrates in a concerted manner over long distances. That said, they are involved in very important cellular processes such as translation initiation, and restarting stalled replication forks (69, 70). The rings formed by the members of SF3—SF6 topologically link the helicases to the DNA substrate. The linkage requires the ring to open or encounter strand breaks for the helicase oligomer to dissociate from its substrate, and the rings are generally made of larger oligomers than those observed in SF1 & SF2 which provides additional contacts for maintaining the interaction between the helicase and its substrate (71). The increase in processivity is important in processes which must occur rapidly such as DNA replication

during cell growth and division.

SF 1 & 2 are the largest superfamilies, and are defined by at least seven motifs in a conserved core region (reviewed in (62). Differences in the amino acid compositions of these motifs, and the substrates of the members distinguish the two families from each other (reviewed in (66, 68)). The core region is responsible for ATP-binding and hydrolysis, as well as nucleic acid binding and translocation (reviewed in (62, 66, 68)). Outside of the core region, these helicases have N-terminal and C-terminal regions that provide a great deal of functional diversity to members of SF 1 & 2.

2.2.2 DEAD-box helicases

The DEAD-box proteins are the largest family within SF2. The family is primarily defined by having the D-E-A-D sequence in motif II, with some variation at the location held by the alanine residue, as well as conservation of several other motifs with a variation distinct from other SF2 families. In addition to its size, the DEAD-box family is also noted for the diversity of cellular processes that employ members of the helicase family. A great deal of this diversity is made possible by the N-terminal and C-terminal domains outside the helicase core (reviewed in (67, 68, 72)).

The RNA helicase members of this family have been the most extensively characterized. The work on the RNA helicases has provided a great deal of understanding of how the DEAD-box proteins function. One notable finding is that some of the RNA DEAD-box proteins

function more like RNA chaperones which bind to duplexed RNA to melt the duplex, and then release their substrate, than proper helicases or translocases which move along length of the nucleic acid strands (73). However, given the relatively small number of DNA DEAD-box helicases which have been studied, it is not clear to what extent some of these behaviors are shared across nucleic acid substrates. I mention all of this again, because based upon a bioinformatic analysis ComFA looks to be a DEAD-box family member (Figures 2.1 & 2.2). Multiple sequence alignments with DEAD-box familiy members, and motif search analyses show the presence of at least five of the seven motifs used to define the DEAD-box helicase family in the primary sequence of ComFA (Figure 2.2) (58). If ComFA does appear to be a member of this family it would confirm that an non-processive helicase is likely the cytosolic driver of processive DNA uptake. Having a non-processive helicase as a major contributor to a processive process makes a clear and interesting paradox, which would suggest that there are methods for modulating or augmenting the processivity, or of creating an apparently processive process from non-processive parts.

2.2.3 ComFA

As previously mentioned in Chapter 1, ComFA is a late *com* protein, part of a large protein complex involved in DNA uptake and transformation in *B. subtilis*. ComFA was originally identified as a late competence *com* gene in an operon designated *comF* during a transposon mutagenesis screen (74). ComFA appears to be the only cytosolic motor that participates

in the DNA uptake machine.

Some initial analyses of the protein showed that it is part of an operon containing three open reading frames, and two of the members (*comFA* and *comFC*) had significant contributions to transformation (58). ComFA, however, had the largest contribution (58). Bioinformatics indicate that it contained some of the motifs found in members of the DEAD-box family of helicases (Figure 2.1). Londoño-Vallejo and Dubnau created a set of in-frame deletion mutants and motif I (Walker A) mutants to test the requirement for ComFA. They found that the in-frame deletions and the Walker A mutants greatly impaired transformation, causing 1 000–fold decreases in transformation efficiency, and also impeded DNA uptake (49, 58). From the tested mutants and the bioinformatic data, they asserted that ComFA was a likely DEAD-box helicase. Their findings were insufficient to solidify that designation, as the GKT motif of the Walker A motif, being required for nucleotide binding is widespread among ATPases, not just helicases, and they did not test any other conserved sequence motifs in ComFA.

Figure 2.1: Schematic of ComFA - Alignment of ComFA and the consensus motif sequences for DEAD-box family helicases. The silver region is the putative zinc finger (see Chapter 3), the red region corresponds to a possible ABC signature linker motif, and the green regions denote DEAD-box helicase motifs. Roman numerals listed above the sequences denote the number motifs. The magnified region shows the region removed by the *comFA*$^{\Delta S1}$ in-frame deletion. The first and last residues of the magnified region are numbered. *Residues mutated in this work.

Figure 2.2: ComFA resembles a DEAD-box protein - Multiple sequence alignment of ComFA with representative DEAD-box proteins performed using M-Coffee (75, 76, 77, 78). Bracketed numbers indicate intervening residues that were removed to condense the sequences. Underlines indicate conserved SF2 DEAD-box helicase motifs. The sequences were obtained from GenBank and were originally described in: DDX1: (79), DeaD: (65), eIF4A: (80), SrmB: (81), UvrB: (82), PRH1: (83), PriA: Foulger: (84), Vasa: (85).

I sought to examine the other DEAD-box motifs present in ComFA to determine if it did fit well as a DEAD-box family member. Having a solid understanding of how the protein likely behaves is very important as we determine how the machinery functions, and develop our understanding of how ComFA is involved in the transformation process.

2.3 Results

2.3.1 ComFA is a DEAD-box helicase

I examined the DEAD-box helicase motifs in ComFA using a transformation efficiency assay. The transformation efficiency assay allowed me to examine the effects of mutations in ComFA on the number of transformants that arise in a population of cells. I designed mutations in the specific nucleotides corresponding to the amino acid residues of interest and introduced those mutations into the *B. subtilis* genome. In the transformation efficiency experiments, I grew cells to competence and then assayed for function of ComFA by looking for the acquisition of an antibiotic resistance gene contained on *B. subtilis* genomic DNA provided to the cells.

I chose to conduct the mutational analyses using mutations introduced at the *comF* locus to resolve some of the problems observed when making complementation constructs expressed from the *yvbJ* locus (See Appendix A). Bioinformatics analysis identifies ComFA as a member of the DEAD–box helicase/translocase family (58, 60). However, previous functional

studies have only confirmed that it is a P-loop ATPase (49). I set out to confirm whether the other DEAD-box family motifs are important to ComFA function, by transformation efficiency analysis (Figure 2.3). I found that relative to wildtype, predicted loss-of-function mutations in motifs I, II, III, and VI led to 100–fold decreases in transformation efficiency. One exception was the $comFA^{S264A}$ mutation, which did not result in a transformation efficiency defect. It has been observed in other DEAD-box helicases generally that either the serine or the threonine in motif III are required for function, but not necessarily both, and some studies actually mutate both residues when analyzing these helicases (86). I also noticed that the $comFA^{\Delta S1}$ in-frame deletion causes a much stronger defect than any of the motif mutations.

2.4 Conclusions

The goal of the work presented in this chapter was to perform a functional analysis of the SF2 motifs present in ComFA. The transformation efficiency analysis demonstrates that each of the four motifs tested are required for efficient transformation in *B. subtilis*. Based on the requirement for the motifs, ComFA clearly is a DEAD-box family helicase.

ComFA contains several of the important motifs to allow its classification, however, the precision of the classification is a bit limited due to the limited breadth of data on similar helicases in bacteria. Most of the DEAD-box helicases that have been studied extensively are RNA helicases found in eukaryotes (66, 87, 88, 89). The relatively recent re-examination of

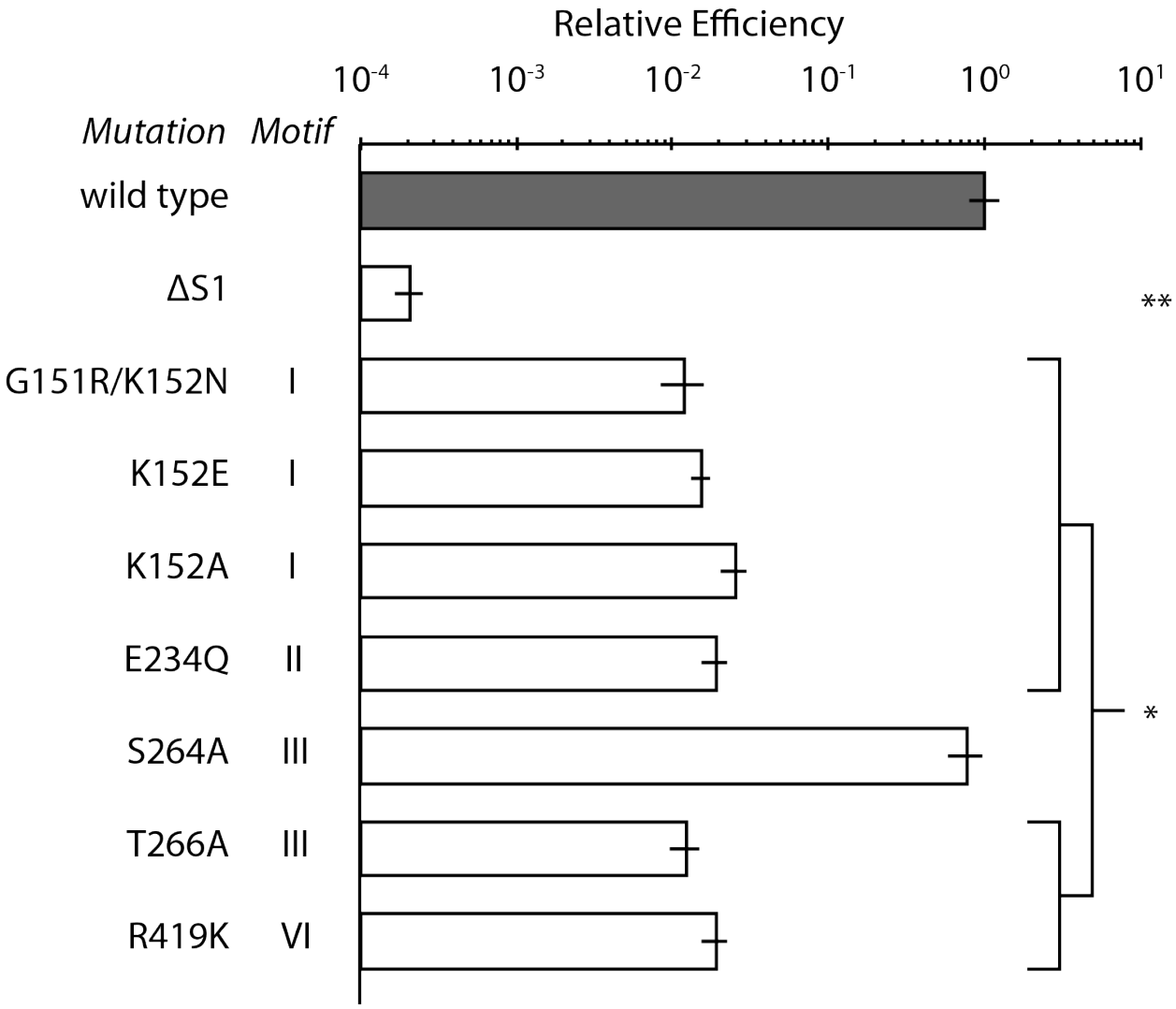

Figure 2.3: DEAD-box motifs are required for ComFA function - Transformation efficiency for mutations of ComFA canonical DEAD-box helicase motifs. All efficiency rates are normalized to wildtype (WT). The ΔS1 strain is included as a *comFA*$^{-}$ control (49, 58). Limit of detection for the assay is 0.5 transformants per CFU per μg of genomic DNA. The relative efficiency axis is a $\log_{10}$ scale. Error bars are standard error, WT n = 35, all mutants n = 5. *p <0.001, **p <0.0001

DEAD-box RNA helicases as RNA chaperones rather than helicases or translocases highlights how the dearth of information can greatly effect how to interpret the classification of ComFA in our models of DNA uptake and transformation (73).

There are two prevailing models for how ComFA may function as part of the machinery: the Brownian ratchet, and the active transport model (35). In the Brownian ratchet model, ComFA introduces directionality to random molecular motion by blocking DNA from exiting back out of the ComEC channel once it has entered the cell. This model is plausible given for example the activity of some DEAD-box RNA helicases, such as Dbp4A which function more as RNA chaperones, and bind and release with ATP-hydrolysis cycles (73). This model, however, restricts the import driving force to the ComG pseudopilis and the proton motive force which are both upstream of the primary DNA binding protein in the system, ComEA (See Figure 1.4). The sequence of components makes this unlikely given the speed of uptake and transformation. The Brownian ratchet may, however, have a role in transformation via contribution from single stranded DNA binding proteins, and could account for the transformation we observe in the absence of a functional helicase protein, and may also provide some insight into the activity differences observed between the helicase motif mutants and the *comFA*$^{\Delta S1}$ in-frame deletion (Figure 2.3). Furthermore, ComFA is present in competence-capable Gm^+ bacterial species and absent in competence-capable Gm^- bacerial species. This leaves the question of why Gm^+ bacteria use a helicase for a process that Gm^- bacteria seem to accomplish using single-stranded DNA (ssDNA) binding

proteins.

We see a decreased level of transformation efficiency in the DEAD-box motif mutants, and a residual level of transformation in the *comFA*$^{\Delta S1}$ strain. From these observations, and the presence of recombination proteins it appears that ComFA has an active role in DNA transport. I would argue that the simplest Brownian ratchet model would cover the mutant phenotypes where essentially the ComFA mutants and the *rec* proteins bind the incoming DNA, and prevent it from flowing back out of the ComEC pore. The functional ComFA, however, may function as a helicase that facilitates uptake by moving along the incoming DNA strand. This facilitation may occur regardless of whether ComFA actively melts the duplex, or steps into space between the strands created from thermal breathing of the DNA double helix. The active molecule can both prevent the back flow of the substrate, and also produce a tension to import the DNA.

Developing a full picture of how ComFA functions during uptake is essential to determining whether a Brownian ratchet is sufficient to describe uptake, or if something more complex is happening. The details for illustrating the behavior will be found in an *in vitro* biochemical analysis of ComFA. A quantitative analysis of the biochemical activities will allow simulations to test models of the uptake system in a variety of conditions and compare the model outputs to reports of uptake lengths and rate (33, 90, 91).

Recent examination of ComGA has revealed that the cytosolic DNA uptake machinery may not need to keep up with the extracellular machinery to achieve efficient transformation

(48). *B. subtilis* appears to have an ability to protect incoming DNA from extracellular nucleases, prior to internalization of the transforming DNA (46, 47, 48). To determine how fast we would expect ComFA to operate, we need more information regarding the capacity of the membrane-proximal space for incoming DNA. The capacity would determine the DNA uptake system's tolerance to asynchrony in the DNA uptake steps. Additional experimentation will be required to determine the contribution to internalization from ComFA, as mutations in ComGA are sufficient to block the internalization process.

Earlier work performed by Londoño-Vallejo and Dubnau used *comFA*$^{\Delta S1}$ as a *comFA*$^{-}$ allele, and showed that the *comFA* motif I mutants were equivalent to the *comFA* null allele (49, 58). It is clear from the *comFA*$^{\Delta S1}$ phenotype in this analysis that the DEAD-box helicase motifs do not account for all of ComFA activity during transformation. We would expect null alleles to be equivalent if they remove all activity attributed to a protein. Here, we see an additional transformation efficiency defect in the *comFA*$^{\Delta S1}$ which suggests that the helicase-dead mutants retain some activity which is then lost when the 113–amino acid region is removed in the ΔS1 mutation. In the next chapter I will investigate one motif that is present in the region removed by the ΔS1 deletion (Figure 2.1).

2.5 Materials and methods

2.5.1 Strains and growth conditions

All *B. subtilis strains* were derived from the prototrophic strain PY79 (92). *B. subtilis* were grown in Luria-Bertani (LB) broth or on LB plates fortified with 1.5 % Bacto agar at 24 °C or 37 °C as appropriate. 10x modified competence (MC) medium was made as described in (93). Competent cells were grown in 1x MC supplemented with 0.3 % (v/v) 1 M $MgSO_4$. When appropriate, antibiotics were included at the following concentrations: 5 μg/ml chloramphenicol (Cm_5), 100 μg/ml spectinomycin ($Spec_{100}$), and 1 μg/ml erythromycin plus 25 μg/ml lincomycin (*mls*).

2.5.2 Plasmid construction

Plasmids used in this work are listed in Table 2.2. Oligonucleotides used in this work are listed in Table 2.3. All isothermal assembly (ITA) reactions were performed as described in (94).

pSC010 [gfp^{mut2b}] was generated by site-directed mutagenesis using oSC012 to add a *Bam*HI restriction endonuclease site to pKL147 (95).

pSC012 [*comFA-gfp*^{mut2b}] was generated in a two-way ligation with an *Eco*RI-*Xho*I PCR product containing P_{comF}-*comFA*. The PCR product was generated by a linking PCR reaction (96) using oSC014, oSC015, oSC016, and oSC017, a small PCR product containing P_{comF}, amplified from *B. subtilis* genomic DNA using oSC014 and oSC015, and a PCR product containing the *comFA* coding sequence amplified from pBB031 using oSC016 and oSC017. The oSC016 oligonucleotide introduces a G to A transition in the start codon for the *comFA* coding sequence, and creates an *Nde*I restriction site.

pSC026 [h_6-*comFA*$^{G151R/K152N}$] was generated by site-directed mutagenesis of pBB031 using oSC069.

pSC036 [*yvbJ::comFA (erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI PCR product amplified from pSC012 using oSC014 and oSC061 and pBB268 cut with *Eco*RI and *Bam*HI.

pSC048 [P_{comF}-*comFA*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-*comFA* from pSC036 into pBlueScriptSK(+) cut with *Eco*RI and *Bam*HI.

pSC051 [P_{comF}-*comFA*^{T266A}] was generated in a two-way ligation with an *Nde*I-*Bam*HI

PCR product amplified from pSC005 using oSC044 and oSC061 into pSC048 cut with *Nde*I and *Bam*HI.

pSC056 [P_{comF}-$comFA^{K152E}$] was generated in a two-way ligation with an *Nde*I-*Bam*HI PCR product amplified from pSC024 using oSC044 and oSC061 into pSC048 cut with *Nde*I and *Bam*HI.

pSC058 [P_{comF}-$comFA^{G151R/K152N}$] was generated in a two-way ligation with an *Nde*I-*Bam*HI PCR product amplified from pSC026 using oSC044 and oSC061 into pSC048 cut with *Nde*I and *Bam*HI.

pSC067 [P_{comF}-$comFA^{S264A}$] was generated in a two-way ligation with an *Nde*I-*Bam*HI PCR product amplified from pSC004 using oSC044 and oSC061 into pSC048 cut with *Nde*I and *Bam*HI.

pSC068 [P_{comF}-$comFA^{K152A}$] was generated by site directed mutagenesis of pSC048 using oSC018.

pSC069 [P_{comF}-$comFA^{E234Q}$]was generated by site directed mutagenesis of pSC048 using oSC008.

pSC073 [P_{comF}-$comFA^{\Delta S1}$] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-*comFA* from pSC048 into pUC19 cut with *Eco*RI and *Bam*HI. The resulting vector was cut with *Sac*I, separated from the fragment produced and re-ligated to produce the $comFA^{\Delta S1}$.

pSC076 [*yvbJ*::P_{comF}-$comFA^{K152A}$] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{K152A}$ from pSC068 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC081 [$comFA^{\Delta S1}$] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{\Delta S1}$ from pSC073 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC104 [*comFA::cat*] was generated by multiple steps involving two-way ligations. The region upstream of *comFA* was inserted via two-way ligation with an *Eag*I-*Sal*I PCR product amplified from *B. subtilis* genomic DNA using oSC032 and oSC033 and pBB028 cut with *Eag*I and *Sal*I. The region downstream of *comFC* was inserted by two-way ligation with a *Sph*I-*Xba*I PCR product amplified from *B. subtilis* PY79 genomic DNA using oSC193 and oSC194 and the plasmid created in the previous two-way ligation cut with *Sph*I and *Xba*I.

pSC216 [P_{comF}-$comFA^{T266A}$ in pMiniMAD2] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{T266A}$ from pSC051 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC217 [P_{comF}-$comFA^{K152E}$ in pMiniMAD2] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{K152E}$ from pSC056 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC218 [P_{comF}-$comFA^{E234Q}$ in pMiniMAD2] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{E234Q}$ from pSC069 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC228 [P_{comF}-$comFA^{G151R/K152N}$ in pMiniMAD2] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{G151R/K152N}$ from pSC058 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC229 [P_{comF}-$comFA^{S264A}$] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{S264A}$ from pSC067 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC230 [P_{comF}-$comFA^{K152A}$] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{K152A}$ from pSC076 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC237 [$comFA^{218\text{-}463}$] was generated by isothermal assembly of a PCR product containing $comFA^{218\text{-}463}$ and 472 bases immediately downstream amplified from *B. subtilis* genomic DNA using oSC311 and oSC312, and pMiniMAD2 cut with *Sma*I.

pSC239 [$comFA^{218\text{-}463,R419K}$]was generated by site-directed mutagenesis of pSC237 using oSC024.

pSC240 [$comFA^{\Delta S1,1\text{-}146}$ in pMiniMAD2] was generated by isothermal (ITA) assembly of a PCR product containing $comFA^{1\text{-}8}$ and 424 bases upstream amplified from *B. subtilis* genomic DNA using oSC313 and oSC335, a PCR product containing $comFA^{\Delta S1,1\text{-}146}$ amplified from pSC081 using oSC314 and oSC334, and pMiniMAD2 cut with *Sma*I.

2.5.3 Construction of *B. subtilis* strains

2.5.3.1 *comF* expression strains

Unmarked mutations were introduced into the genome as performed in (97) with the following modifications. The transduction step was omitted. Transformed colonies carrying the

pMiniMAD2 constructs were re-streaked on *mls resistance* selective plates, and grown at 37 °C overnight. One colony from 5—8 of the re-streaked isolates were used to inoculate a 5 ml LB culture, and grown rolling at 37 °C for 6–8 hours. Cells were diluted 1:1 000 in 25 ml LB and grown at the permissive temperature for 24 hours, diluted again, and grown up at the permissive temperature for an additional 24 hours. Mutations were verified by sequencing of PCR products amplified from purified genomic DNA of *mls-sensitive* isolates or directly from cells. PCR products were generated using oSC085 and oSC086.

2.5.3.2 Genomic DNA preparation and PCR amplification

Patches from the desired strains were used to inoculate 3—6 ml LB liquid cultures. Cultures were grown up at 37 °C for 4 hours, or 24 °C overnight. To make the genomic DNA preparations, 1 ml of culture was pelleted at 21 130 x g in a microcentrifuge for 1 min. Growth media was removed and the cells were resuspended in 500 μl genomic DNA lysis buffer (20 mM Tris-HCl pH 7.5, 50 mM EDTA, 100 mM NaCl). To breakdown the cell wall, lysozyme was added to the suspension to a concentration of 2 mg/ml and the cells were incubated at 37 °C for 20 minutes. To lyse the cells 30 μl of 20 % (w/v) *N*-Lauroylsarcosine was added to suspension, and mixed by vortexing. DNA extracted using a phenol treatment step, followed by a phenol chloroform treatment step. The aqueous phases were harvested for each step. To the aqueous phase from the second step 3 M sodium acetate, pH 5.2 was added to 10 % (v/v) and mixed. To precipitate the DNA 2x the solution volume of 200 proof

ethanol was added to the solution, and mixed. The DNA was precipitated by centrifugation at 21 130 x *g*. The pellet was washed once with 70 % ethanol and then dried. The dried pellet was resuspended in 500 μl TE buffer.

2.5.3.3 Colony PCR

PCR products for sequencing were generated from *B. subtilis* cells in the following manner. Cells from a patch on an LB/agar plate were resuspended in 20 μl of 0.05 mg/ml lysozyme solution. Cells were heated in a thermocycler to 37 °C for 15 minutes, then to 99 °C for 10 minutes, and then cooled to 20 °C. PCR reactions were conducted using Phusion Polymerase (ThermoFisher Scientific, Inc.) in 20 μl reactions. For each reaction 1 μl of the *B. subtilis* preparations were used as the DNA template. The primers oSC085 and oSC086 were used for amplification of the target region.

2.5.4 Transformation efficiency analysis

2.5.4.1 *comF* expression strains

A fresh colony was picked from an LB plate grown overnight at 37 °C was used to inoculate 5—6 ml LB in 18 mm glass tube. Culture was grown rolling at 24 °C for 12—16 hours. Cells from cultures with a final OD_{600} of 0.2—1.2 were harvested by pelleting at 6 010 x *g*, washed 3x with 1 ml 1x PBS, and resuspended in 500 μl of 1x MC. The OD_{600} was measured following resusupension in 1x MC. The washed cells were used to inoculate a 1 ml 1x MC

culture to a starting OD_{600} of 0.01. Cells were grown rolling at 37 °C for 5 hours. At 5 hours post-inoculation 900 μl the culture was transferred to 13 mm glass tubes, 0.9 μg gSC018 genomic DNA was added, and the $MgSO_4$ concentration was increased to 8 mM. Cells were grown rolling at 37 °C for 2 additional hours. Serial dilutions of the culture were made in 1x PBS. For the CFU counts, 100 μl of the 10^{-6} dilution was plated in duplicate on non-selective media. Dilutions to allow for 50—1 000 CFU per plate, when possible, were plated on selective plates containing 100 μg/ml spectiniomycin. Plates were incubated at 37 °C overnight, and colonies counted the following day. Transformation efficiency was calculated as

$$\eta_s = \frac{C_r}{C_T \rho_D} \tag{2.1}$$

where C_r = average resistant CFU (transformants), C_T= average total CFU, and ρ_D = DNA concentration in μg/ml. Each round was normalized to wildtype run at the same time. Limit of detection for the assay is 0.5 transformants per CFU per μg of genomic DNA.

2.5.5 Statistical analysis

Transformation efficiency statistical analysis was conducted in R. Relative efficiencies from experiments were used for the analysis. Relative efficiency data for wild type was generated from relative efficiency of a given experiment relative to the average wild type efficiency to

from the entire dataset. Data was transformed by

$$\eta^*_{i,w} = \frac{180 sin^{-1}\left(\sqrt{\left(\frac{\eta_{i,w}}{6}\right)}\right)}{\pi} \tag{2.2}$$

where $\eta_{i,w}$ is the relative efficiency of a strain compared to wild type, prior to analysis of variance (ANOVA). For wildtype, the relative efficiency was determined for each experiment compared to the average wild type efficiency over all experiments to be compared. Following ANOVA, the Dunnett's multiple comparison test was applied post-hoc to determine if the mean relative transformation efficiencies of the mutant strains were significantly different from wild type.

Table 2.1: Strains used in Chapter 2

Strain	Genotype	Reference
B. subtilis PY79		(92)
bSC007	*comFA::Tn524 (erm)*	This work
bSC016	*comFA::cat*	This work
bSC017	*yvbJ::*P_{comF}*-comFA*^{T266A} *(erm)*	This work
bSC018	*yvbJ::*P_{comF}*-comFA*^{R419K} *(erm)*	This work
bSC022	*yvbJ::*P_{comF}*-comFA*^{K152E} *(erm)*	This work
bSC023	*yvbJ::*P_{comF}*-comFA*$^{G151R/K152N}$ *(erm)*	This work
bSC025	*yvbJ::*P_{comF}*-comFA*^{S264A} *(erm)*	This work
bSC026	*yvbJ::*P_{comF}*-comFA*^{K152A} *(erm)*	This work
bSC027	*yvbJ::*P_{comF}*-comFA*^{E234Q} *(erm)*	This work
bSC031	*yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$ *(erm)*	This work
bSC032	*yvbJ::*P_{comF}*-comFA (erm)*	This work
bSC042	*yvbJ::cat*	This work
bSC049	*comF::cat*	This work
bSC180	*comFA*^{T266A}	This work
bSC182	*comFA*^{K152E}	This work
bSC183	*comFA*^{E234Q}	This work

Continued on next page

Table 2.1 – *Continued from previous page*

Strain	Genotype	Reference
bSC188	*comFA*$^{G151R/K152N}$	This work
bSC190	*comFA*^{S264A}	This work
bSC192	*comFA*^{K152A}	This work
bSC199	*comFA*^{R419K}	This work
bSC201	*comFA*$^{\Delta S1}$	This work
E. coli DH5α	F$^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r$_K^-$ m$_K^+$), λ–	

Table 2.2: Plasmids used in Chapter 2

Name	Genotype or characteristics	Reference
pBB028	*cat bla*	B.M. Burton
pBB031	P_{T7}-h_6-*comFA kan*, derived from pET28b(+)	B.M. Burton
pBB268	*yvbJ::erm bla*	Gift from D. Rudner
pBB278	*yhdGH::spec bla*	B.M. Burton
pBlueScriptSK(+)	*bla*	
pET28b(+)	*kan*	Novagen
pKL147	*gfp*^{mut2b}	(95)
pMiniMAD2	*ori*^{BsTs} *bla erm*	(97)
pSC010	*gfp*^{mut2b}, derived from pKL147	This work
pSC012	P_{comF}-*comFA-gfp*^{mut2b}, derived from pSC010	This work
pSC026	*comFA*$^{G151R/K152N}$, derived from pBB031	This work
pSC036	*yvbJ::*P_{comF}*-comFA (erm)*	This work
pSC048	pBlueScriptSK(+) with P_{comF}-*comFA bla*	This work
pSC051	*comFA*^{T266A}, derived from pSC048	This work
pSC056	*comFA*^{K152E}, derived from pSC048	This work
pSC058	*comFA*$^{G151R/K152N}$, derived from pBlueScriptSK(+) and *comFA*$^{G151R/K152N}$ from pSC026	This work
pSC067	*comFA*^{S264A}, derived from pSC048	This work
pSC068	*comFA*^{K152A}, derived from pSC048	This work
pSC069	*comFA*^{E234Q}, derived from pSC048	This work

Continued on next page

Table 2.2 – *Continued from previous page*

Strain	Genotype	Reference
pSC073	*comFA*$^{\Delta S1}$, derived from *Sac*I digest of P_{comF}-*comFA* inserted into pUC19	This work
pSC076	pBB268 with insert from pSC068	This work
pSC081	pBB268 with insert from pSC073	This work
pSC104	*comF::cat bla*	This work
pSC216	pMiniMAD2 with insert from pSC051	This work
pSC217	pMiniMAD2 with insert from pSC056	This work
pSC218	pMiniMAD2 with insert from pSC069	This work
pSC228	pMiniMAD2 with insert from pSC058	This work
pSC229	pMiniMAD2 with insert from pSC067	This work
pSC230	pMiniMAD2 with insert from pSC076	This work
pSC237	pMiniMAD2 with *comFA*$^{218\text{-}463}$ and 472 bases downstream	This work
pSC239	*comFA*^{R419K}, derived from pSC237	This work
pSC240	pMiniMAD2 with insert derived from pSC081	This work
pUC19	*bla*	

Table 2.3: Oligonucleotides used in Chapter 2

Name	Sequence
oSC002	CTGGGCGGTTTGCGGCGCTGGC**GCT**ACA GAAATGCTGTTTCCTGGTATA
oSC008	GATGCAATCGATGTTATGATCATTGAT**C**A **G**GTTGACGCTTTTCCATATTCTGC
oSC009	CAGCACCCTCGTTTATTTA**GCG**GCAACAC CTCCTAAAGAATT
oSC010	CACCCTCGTTTATTTAAGTGCA**G**C**G**CCTC CTAAAGAATTAAAAAGAAAAGC
oSC011	CACCCTCGTTTATTTAAGTGCA**G**C**G**CCTC CTAAAGAATTAAAAAGAAAAGC
oSC014	G<u>GAATTC</u>CAAATCTCCGTTTTTAGAGCGG AGATTTTTTTATATTCTTA
oSC015	CCG<u>CTCGAG</u>CG<u>GAATTCCATATG</u>GCACGC CTCCTTTCGAAACAGTATG

Continued on next page

Table 2.3 – *Continued from previous page*

Name	Sequence[a,b]
oSC016	GGAATTCCATATGAATGTGCCAGTTGAAA AAAACAG
oSC017	CCGCTCGAGGTCTGTACATTCAACTTTTG CTGCC
oSC018	GCGGTTTGCGGCGCTGGCGAAACAGAAA TGCTGTTTCCTGGTATAGAATC
oSC024	CATATTCTTTATGCCGGCCGGTT**T**TTCCT GCAATTTGAACAAGTGCGCT
oSC032	CATGATCGGCCGCTTAAAGCTGCTGAATT AATCAAAAACGGAGC
oSC033	TTACGCGTCGACTAAAAAAATCTCCGCTC TAAAAACGGAGATTTG
oSC044	GTTTACTTTAAGAAGGAGATATACCATGG GCAGCAGCC
oSC061	CGGGATCCTAGTCTGTACATTCAACTTTT GCTGCC
oSC069	CTGGGCGGTTTGCGGCGCT**C**GCAA**C**ACA GAAATGCTGTTTCCTGGTATAGAATC
oSC085	TGTATCCATTTGACTCAGAGATCAGC
oSC086	TCTCTGATCCTTGTTCTCCACACC
oSC193	ACATGCATGCATGATTCTGTTTTTATGCC GATATAATC
oSC194	GCTCTAGAGTTGCAGTCTTTAAACAATCT TAACCC
oSC311	CAGGTCGACTCTAGAGGATCCCCCAGCTT TTGCGATATAAAGATGCAATC
oSC312	GTGAATTCGAGCTCGGTACCCGATTTTCT TTAATTTGCTTCTGCAAGAATAAC
oSC313	CAGGTCGACTCTAGAGGATCCCCCAAGCC TTCATTGGTAGTCTTCTAAAGGTAAAG
oSC314	GTGAATTCGAGCTCGGTACCCGGTGCTGT TTTTCTTTCTTGCTTTTTG
oSC334	GTTTCGAAAGGAGGCGTGCTAT**G**TGAAT GTGCCAGTTGAAAAAAACAG
oSC335	CTGTTTTTTTCAACTGGCACATTCACATA GCACGCCTCCTTTCGAAAC

Continued on next page

Table 2.3 – *Continued from previous page*

Name	Sequence[a,b]

[a]Bold-face indicates mutagenic residues.
[b]Underlines indicate restriction sites.

3

Chapter 3: Characterization of a zinc finger in ComFA

3.1 Abstract

Overall our understanding of how ComFA functions has been limited to how it functions as a DEAD-box helicase. However, it appears that ComFA may have some additional features that affect its function. Here, I have identified a new metal-binding motif in ComFA that is important to its function. The metal-binding motif is required for zinc binding, and is independent from the helicase function of the protein. While the mechanism of the contribution has yet to determined, its identification is an important addition our understanding of ComFA function.

3.2 Introduction

Metal coordination sites often confer important biochemical and structural features to proteins. These features provide mechanisms for catalysis of chemical reactions, and stabilize tertiary structure elements that mediate intermolecular interactions such as DNA binding and sequence recognition, and protein ligand binding interfaces.

3.2.1 Metal-binding and biochemistry

Metal co-factors are found in many enzymes important to biology. They are often involved in catalyzing reactions such as hydrolysis of amide bonds in metalloprotease substrates, or phosphodiester bonds in nuclease substrates, or even facilitating the hydrolysis of nucleotide triphosphates to utilize the energy stored in the phosphate bonds (98, 99, 100). Metal co-factors are also involved in the structural integrity of some proteins as exemplified by the zinc-finger in RecQ, and transcriptional regulation proteins (101, 102). Specificity of the coordinated metal is also important. For example, the fidelity in DNA polymerase is altered when magnesium is substituted for manganese (103).

3.2.2 Zinc and transition metals

Generally metal cofactors that are biologically active are either alkaline earth metals, such as magnesium and calcium, or transition metals such as zinc and manganese. Due to the

differences in size and valence structures of these metals, how they are coordinated is also different between the two classes (104, 105).

For the purposes of this discussion, I am going to focus on the transition metals, specifically zinc. Zinc is most often found as Zn^{2+} in the biological context. As with many other transition metals it is required for cellular function in small amounts, and too much is actually toxic to cells (reviewed in (106)). Zinc is coordinated in many ways in different proteins, and how it is coordinated generally reflects its function in the associated protein. For example, the zinc fingers most often observed in eukaryotes as part of transcription factors are coordinated in C_2H_2 motifs, and confer sequence specificity to those proteins through the folded structures created to properly coordinate the metal (107, 108, 109, 110). There are also coordination sites that consist entirely of cysteines, such as the C4 zinc finger found in RecQ (101). These C4 zinc fingers are thought to be primarily involved in conferring structural stability to a particular fold rather than being involved directly in catalytic activity. There are also many variations and elaborations on themes, including coordination facilitating dimerization, sites which bind multiple metal ions, and variations in the numbers and types of residues involved in coordinating metal ions (reviewed in (111)).

Several bioinformatic programs have been developed to identify potential zinc finger motifs in proteins, and which amino acid residues likely coordinate the metal (112, 113, 114, 115, 116, 117). The algorithms used vary between programs, and overall their utility is limited in some cases due to the amount of variation in the structures and sequences which

allow the coordination of zinc and other transition metal cations.

From the work I will discuss here, it appears that ComFA has metalloenzymatic activity beyond the expected coordination of magnesium ions to facilitate ATP hydrolysis in the helicase core of the protein. The metal binding appears to be important to its activity, however, it is unclear how the metal binding is required.

3.2.3 The putative zinc finger in ComFA

The initial characterization of ComFA involved a set of in-frame deletions within the coding sequence of the protein. All of the deletions created loss-of-function alleles in the protein, with differing levels of severity (58). All of the in-frame deletions in that work removed regions of the N-terminus of ComFA. The *comFA*$^{\Delta S1}$ mutant was chosen as a *comFA*$^{-}$ allele. Previous work on ComFA showed the Walker A mutants to be equivalent to the *comFA*$^{\Delta S1}$ (49), which suggested that all of the activity for ComFA during transformation is ATPase dependent.

In the work described in the previous chapter, I showed that the helicase motif mutation are not equivalent to *comFA*$^{\Delta S1}$. As the *comFA*$^{\Delta S1}$ allele produces a more severe phenotype (see Figure 2.3), there may be additional factors which contribute to ComFA function. A bioinformatic analysis of ComFA showed that there may be additional motifs amino-terminal to the Walker A motif (will be discussed later), which could could account for the difference (Figure 3.1). Even so, bioinformatics fell short in identifying the new motif and activity I

will discuss in this chapter. Here I will describe the process of characterizing the putative zinc finger in ComFA.

3.3 Results

3.3.1 Discovery of possible metal-binding activity

Discovering that ComFA had a metal-binding activity occurred largely by accident. The new motif and its function as part of ComFA was uncovered while attempting to develop a method to observe the localization of ComFA via fluorescence microscopy. The process of how this was uncovered is discussed at length in Appendix B.

3.3.2 Bioinformatic analysis

When beginning work on ComFA, I subjected the amino acid sequence to several motif searches. The search provided a great deal of information, including DEAD-box helicase motifs examined in Chapter 2, and a putative domain boundary in the protein. However, none of the general, or zinc-finger specific databases called the C4 zinc finger in ComFA (113, 114, 115, 116, 117). The zinc finger was identified manually by examination of the primary sequence of ComFA (Figure 3.1). Interestingly, at least one of the servers used for the predictions identified putative zinc-binding sites in ComFB and ComFC (116, 117).

As the bioinformatic analyses did not identify locations for metal-binding motifs in the

sequence, it was a bit less than straightforward to find the metal-binding motif in ComFA. The first indication of the metal binding motif was the result of cross-reactivity with a biarsenal dye (FlAsH) (see Appendix B). I was able to find a possible C4 zinc finger motif by a manual search when looking for potential mutation sites to possibly block the cross-reactivity with the biarsenal FlAsH dye.

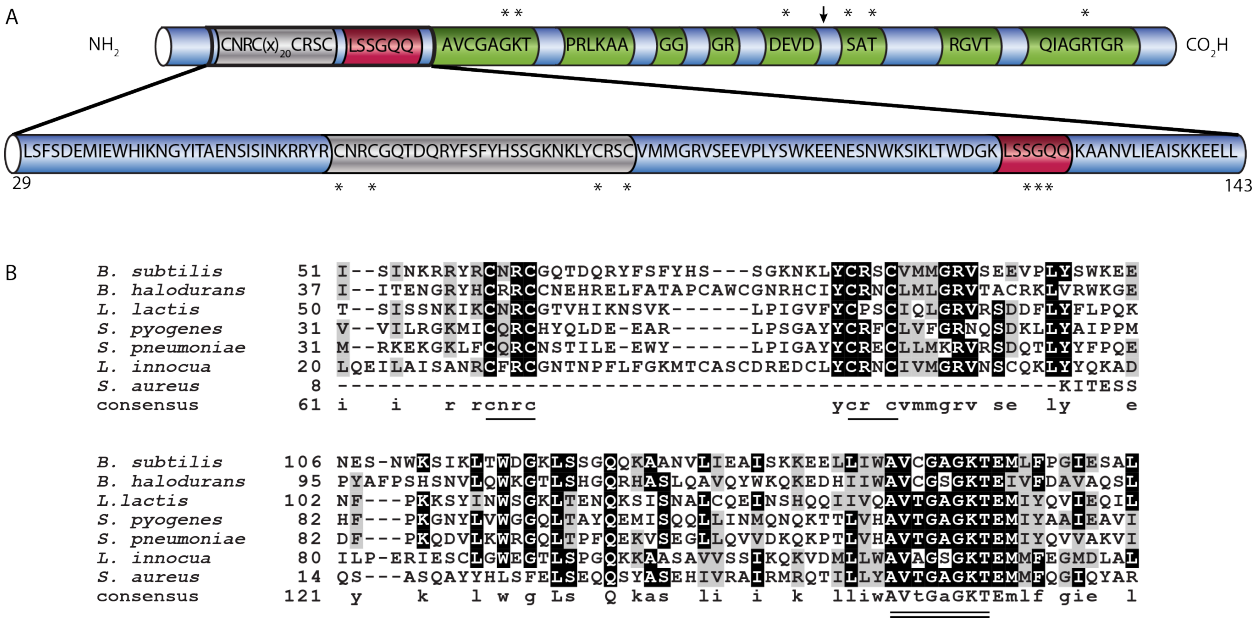

Figure 3.1: ComFA contains a conserved tetracysteine motif - (A) Schematic of the primary amino acid sequence of *B. subtilis* ComFA. The magnified section is the region removed by the *comFA*$^{\Delta S1}$ in-frame deletion. Arrow indicates the C-terminus of the *comFA*$^{1\text{-}252}$ construct. *Residue was mutated in work described. (B) Multiple sequence alignment of ComFA and homologs from other species. Multiple sequence alignment was performed using PSI-Coffee (78, 118) and modified manually to improve the match. Sequences retrieved from GenBank. The sequences selected for each species are as follows with Genbank accession numbers and references: *Bacillus halodurans* [NP_244493, (119)], *Lactococcus lactis* [NP_267246, (120)], *Streptococcus pyogenes* [NP_269668, (121)], *Streptococcus pneumoniae* [NP_346619, (122)], *Listeria innocua* [NP_471986, (123)], *Staphylococcus areus* [YP_005754683, (124)]. Double underline indicates the SF 2 motif I for anchoring reference.

3.3.3 Zinc finger is important for ComFA function

Mutating individual cysteines produced transformation efficiency defects of approximately 10–fold from wildtype (Figure 3.2A). The more amino-proximal cysteines (C60/C63) appear to create slightly more severe defects when mutated than the two amino-distal members (C84/C87) of the motif. Mutating both of the amino distal members or mutating all four of the cysteines create a 100–fold decrease in transformation efficiency, which is equivalent to the defects observed for the DEAD-box helicase motif mutants.

Next we tested the C4 motif. Mutating the individual cysteines in the motif to serines produced a marginal defect in transformation efficiency. The conversion of at least two cysteines to serines was required to produce a transformation efficiency defect similar to what we observe with the canonical DEAD-box mutants (Figure 3.2A). Observing this defect suggested that the zinc finger may hold a role in transformation, and so the next step was to confirm that the motif binds metal.

3.3.4 The zinc-binding activity has an independent contribution to ComFA function from the DEAD-box helicase motifs

Finally, we tested the C4 zinc finger to determine whether its contribution to transformation efficiency was distinct from the contribution we observe from the canonical DEAD-box helicase mutants. To test for a combined defect we made mutants containing the

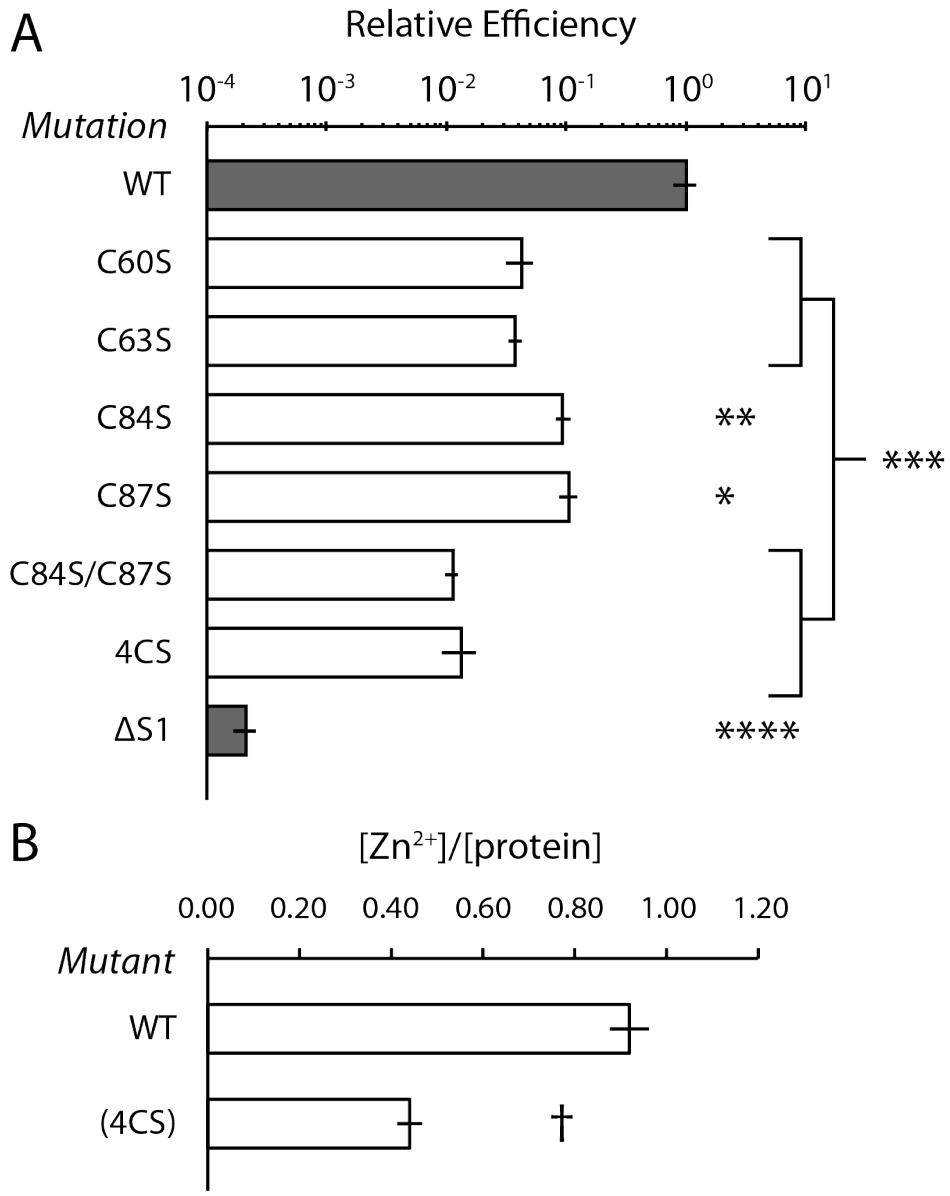

Figure 3.2: Cysteines are required for efficient transformation and zinc binding - (A)Transformation efficiency for mutations of ComFA putative C4-zinc finger motif. All efficiency rates are normalized to wildtype(WT). WT and ΔS1 values are the same as Figure 2.3. Limit of detection for the assay is 0.5 transformants per CFU per µg of genomic DNA. The relative efficiency axis is a $\log_{10}$ scale. Error bars are standard error, WT n = 35, all mutants n = 5. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$(B) Stoichiometry of zinc to protein in WT ComFA and mutants. Ratio determined by PAR A_{500} nm and protein concentration. Error bars are standard error, n = 3. † p = 0.00008.

C60S/C63S/C84S/C87S (4CS) mutations and the each of the canonical DEAD-box motifs previously mutated. The transformation efficiency experiments showed that the C4-zinc finger created an additive, greater than 10–fold defect in transformation efficiency than the DEAD-box mutants that created a 100–fold defect (Figure 3.3).

When we combine the $comFA^{4CS}$ mutation with canonical DEAD-box helicase motif mutants, we see an additional defect in the transformation efficiency of approximately an additional 10–fold decrease, which is comparable to the defect observed in the $comFA^{\Delta S1}$ mutant strain.

3.3.5 Purification of ComFA

Prior to this work, attempts to purify ComFA have been limited to being sufficient for antibody production. The protein generated as an antigen was purified from inclusion bodies (59). Inclusion bodies are large insoluble protein aggregates. Generally the polypeptides are misfolded, and are often solubilized using detergents or chaotropic solutes (125, 126). While the extraction process may not affect antibody generation, it can destroy enzymatic activities. In order to examine the metal-binding activity of the ComFA protein I needed to produce soluble protein which was folded such that any potential metal coordination is produced and maintained.

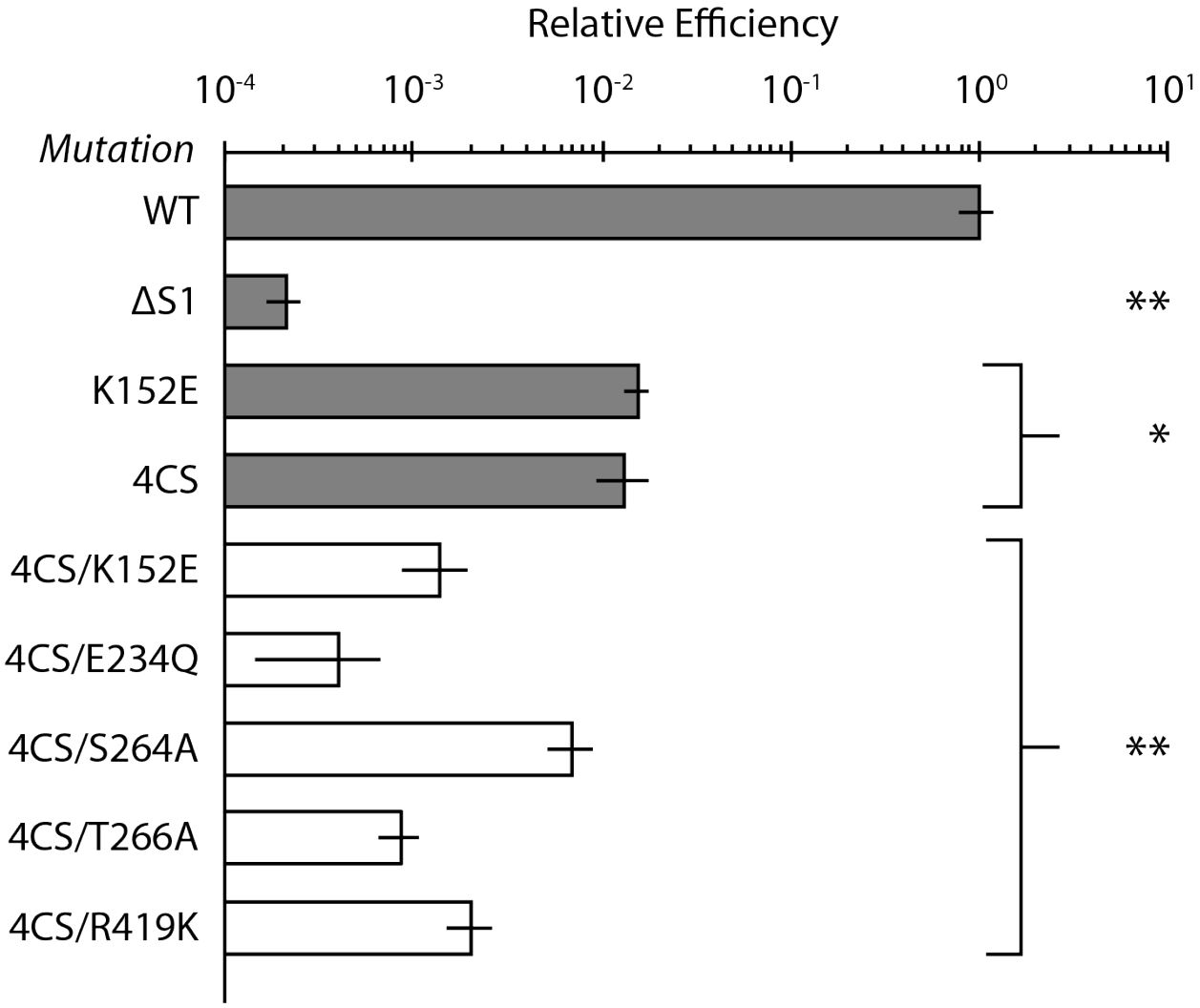

Figure 3.3: The C4-zinc finger requirement is independent of the DEAD-box motif requirement - Transformation efficiency for mutations of ComFA putative C4-zinc finger motif combined with canonical DEAD-box helicase motifs. All efficiency rates are normalized to wildtype (WT). WT and ΔS1 values are the same as Figure 2.3. Limit of detection for the assay is 0.5 transformants per CFU per µg of genomic DNA. The relative efficiency axis is a $\log_{10}$ scale. Error bars are standard error, wildtype n = 35, all mutants n = 5. *p <0.001, **p <0.0001

3.3.5.1 Purification of H_6-ComFA

Previous attempts have generally consisted of adding a hexahistidine (H_6) tag to the amino-terminus in ComFA, and purifying protein expressed in *E. coli* from inclusion bodies by denaturing the protein and purifying the protein by nickel IMAC. While denaturation has worked well for purifying a number of proteins, finding proper refolding conditions can be difficult to find, and refolding often results in a low yield of refolded protein. When beginning this work I started witha H_6-ComFA construct built by B.M. Burton. When expressing the protein in *E. coli* I found that while I had strong induction of the protein, all of the protein was expressed in inclusion bodies (Figure 3.4).

3.3.5.2 Purification of MBP-ComFA

In an attempt to obtain soluble ComFA from *in vitro* expression, I made translational fusion constructs for expression in *E. coli* which added the solubility tag MalE (MBP) to the N-terminus of ComFA. These constructs contained proteolytic cleavage sites to allow separation of the solubility tag from ComFA.

My initial MBP fusion was made by inserting the *comFA* coding sequence into the pMAL-c5E vector available from New England Biolabs. The vector contains a recognition and cleavage sequence for enterokinase. While this vector provided a solubility and expression level improvement over the H_6-ComFA construct, the recognition sequence provided poor specificity for the enterokinase cleavage (Figure 3.5).

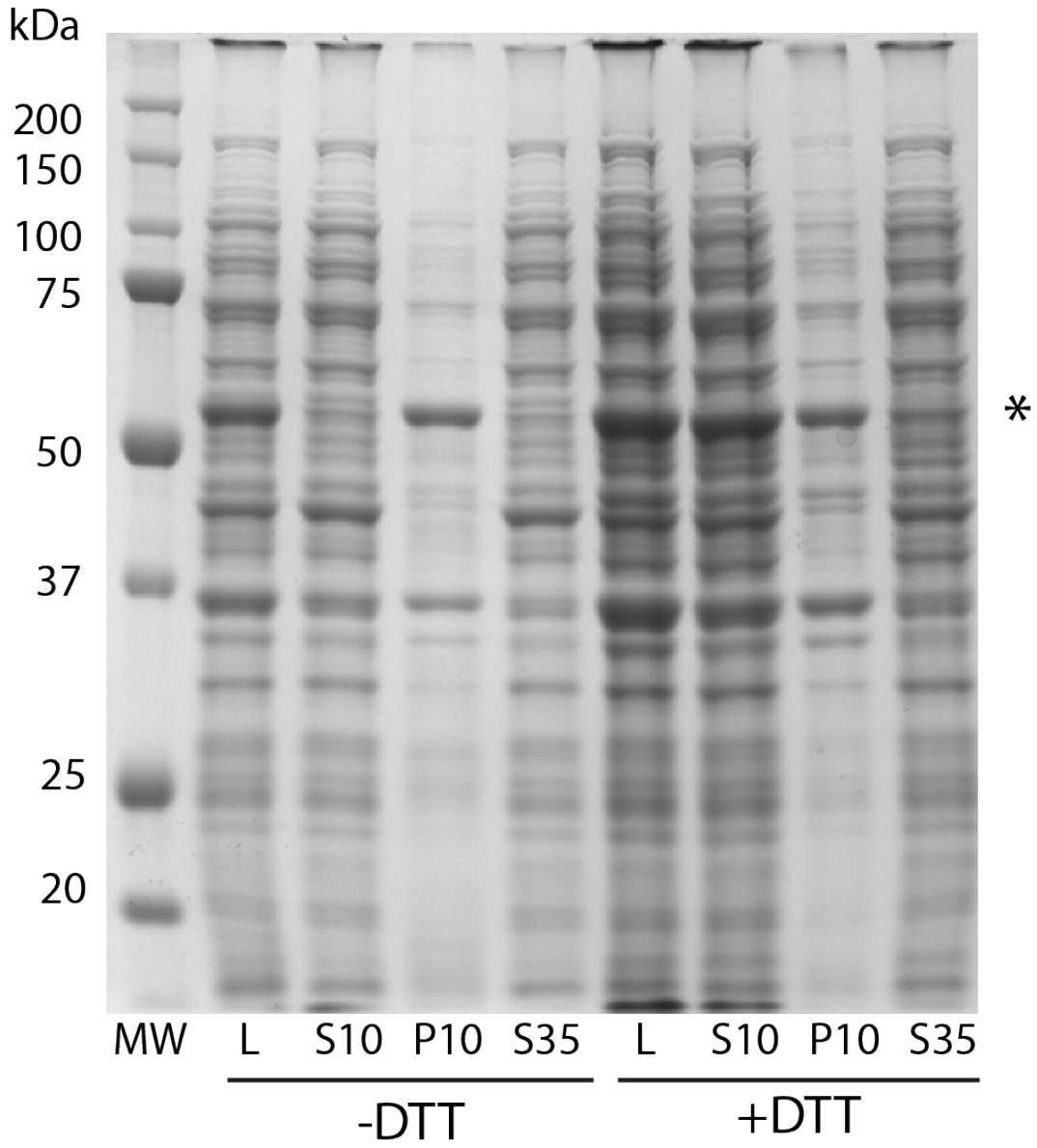

Figure 3.4: H_6-ComFA is insoluble - SDS-PAGE gel of a solubility test for H_6-ComFA. The molecular weight values associated with the molecular weight marker are listed to the left of the gel. MW: Molecular weight marker. L: Cell lysate, S10: supernatant following 10 000 x *g* centrifugation, P10: pellet following 10 000 x *g* centrifugation, S35: supernatant following 100 000 x *g* centrifugation. Solubility was tested in the presence or absence of 1 mM dithiothritol. *Indicates the migration of H_6-ComFA.

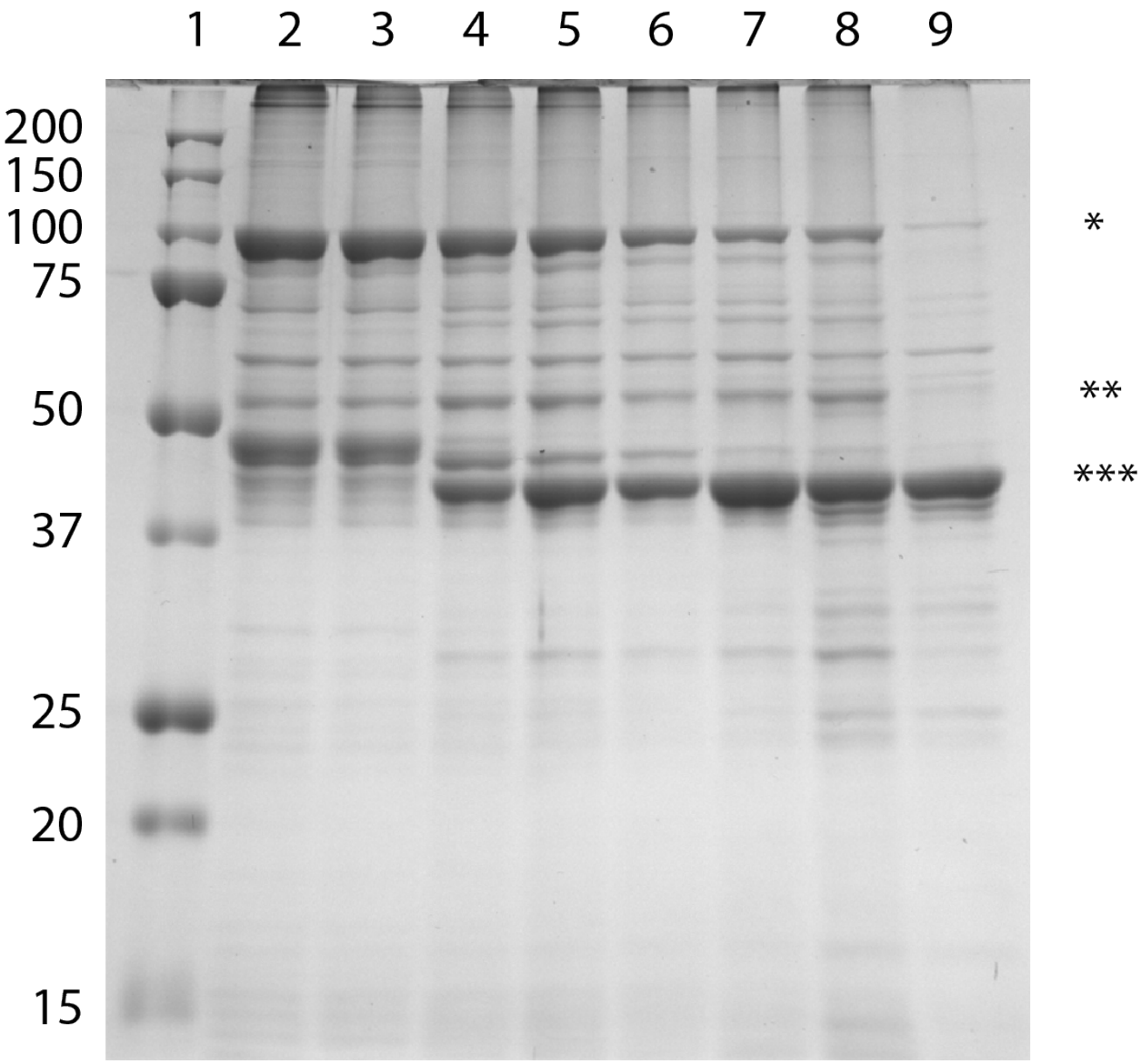

Figure 3.5: MBP-ComFA Enterokinase Cleavage - Cleavage of MBP-ComFA by enterokinase. 10 % Acrylamide SDS-PAGE gel. 1, Precision Plus Protein Standards; 2, -ent 1 hour; 3, -ent 3 hours; 4, 0.0075 % 1 hour; 5, 0.0075 % 3 hours; 6, 0.0225 % 1 hour; 7, 0.0225 % 3 hours; 8, 0.0675 % 1 hour; 9, 0.0675 % 3 hours. Percentages are w/w compared to MBP-ComFA. * indicates MBP-ComFA, ** indicates ComFA, *** indicates free MBP

To resolve the protease cleavage problem I modified the pMAL-c5E vector such that the enterokinase cleavage site was replaced with a 3C human rhinovirus protease recognition sequence, which could be cleaved by PreScission Protease. The new construct (pSC042) performed similarly to the pMAL-c5E fusion construct in solubility and expression. However, it performed much better when cleaved using PreScission Protease expressed and purified following a protocol from W. Weihofen of R. Gaudet's laboratory (Figure 3.6). I used soluble protein purified from *E. coli* using a sepharose dextrin resin, and then cleaved with PreScission protease for the ComFA Zn-IMAC analysis (Figure 3.10).

Following the discovery that ComFA may contain a zinc-binding motif (see Appendix B) I decided to test if the addition of zinc would improve the solubility of the protein expressed in *E. coli*. Work with RecQ has shown that C4-zinc fingers can be important to protein stability (101). To determine whether zinc provided during expression would help with ComFA solubility I made C-terminal truncations near a domain boundary determined by the SCRATCH database (114). ComFA$^{1-252}$ was chosen for additional analysis. In the ComFA$^{1-252}$ the C-terminal half of the protein is removed, including motifs III, V, and VI (Figure 3.1A)When expressed in *E. coli*, the MBP-ComFA$^{1-252}$ fusion showed greatly improved solubility when Zn^{2+} was included in the growth medium during induction (Figure 3.7). Having solved a great deal of the solubility problems with ComFA I decided to test the functional state of the protein by using an electrophoretic mobility shift assay (EMSA). While developing conditions for the assay I found that there was a lot DNA apparently bound

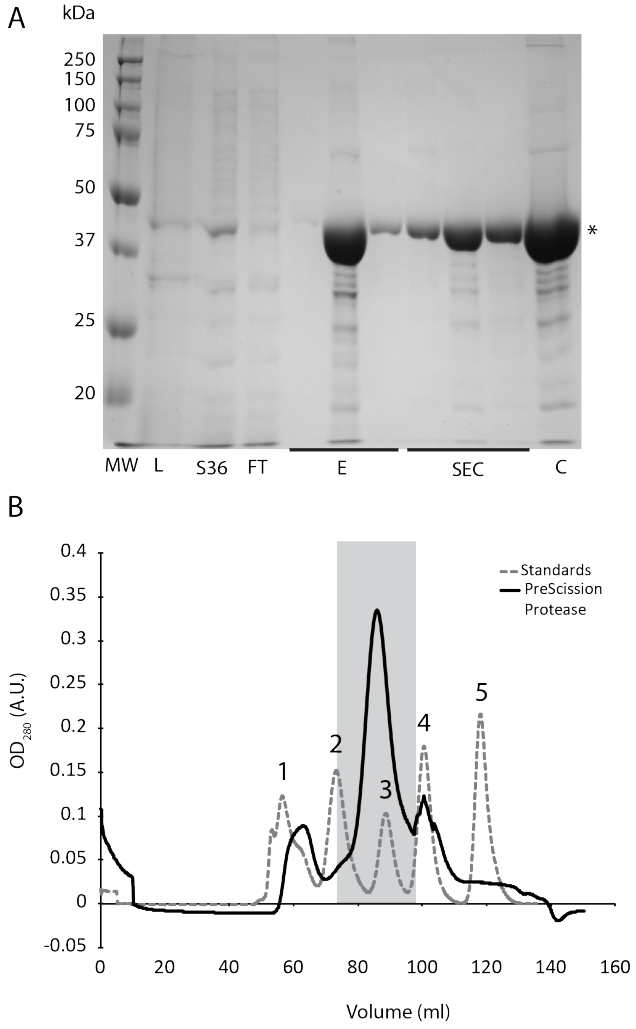

Figure 3.6: Purification of PreScission Protease - (A) SDS-PAGE gel of purified PreScission protease. PreScission Protease expressed in *E. coli* was purified by glutathione-S-transferase (GST) affinity chromatography and then by size exclusion chromatography (SEC). MW: Molecular weight marker, L: Cell lysate, S36: supernatant from 36 Krpm centrifugation, FT: Flowthrough from GSTPrep column, E: Elution from GSTPrep column, SEC: Fractions collected from Superdex 200 column, C: Concentrated SEC fractions *PreScission Protease. (B) Absorbance traces from SEC. Solid black line represents OD_{280} from SEC to purify PreScission Protease. Dotted gray line represents a trace from molecular weight standards on same Sephadex 200 column. 1: bovine thyroglobulin (670 kDa), 2: bovine γ-globulin (158 kDa), 3: chicken ovalbumin (44 kDa), 4: horse myoglobin (17 kDa), 5: Vitamin B_{12} (1.35 kDa). Molecular weight standard trace amplitude was increased to improve visibility on graph.

to the MBP-ComFA protein I had purified, which interfered with the ability to analyze the EMSA results (Figure 3.8A).

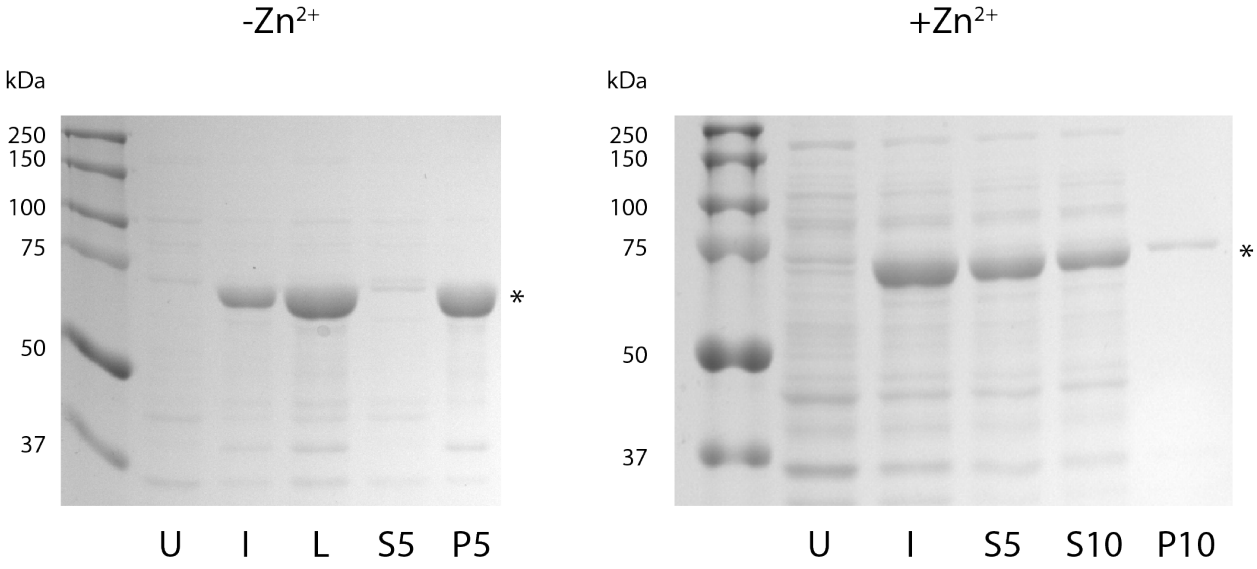

Figure 3.7: Zinc improves the solubility of MBP-ComFA - Coommassie gels from preparations of MBP-ComFA$^{1-252}$. MW: Molecular weight marker. U: Pre-induction. I: Post-induction. L: Lysate. S5: Supernatant from 5 000 x g centrifugation. P5: Pellet from 5 000 x g centrifugation. S10: Supernatant from 10 000 x g centrifugation. P10: Pellet from 10 000 x g centrifugation. * Indicates MBP-ComFA$^{1-252}$

I found that polyethylenimine (PEI) precipitation provided a promising method for removing the DNA bound to the protein (Figure 3.8) as other methods including anion exchange chromatography, and adsorption to hydroxyapatite followed by high salt washes failed to remove significant DNA, and allow recovery of the protein. A PEI precipitation using 0.12 % w/v PEI allows for precipitation of the fusion protein, which can be recovered in 500 mM NaCl. Residual PEI can be removed by sepharose dextrin affinity chromatography, which immobilizes the MBP-ComFA fusion. The precipitation also separates the full-length

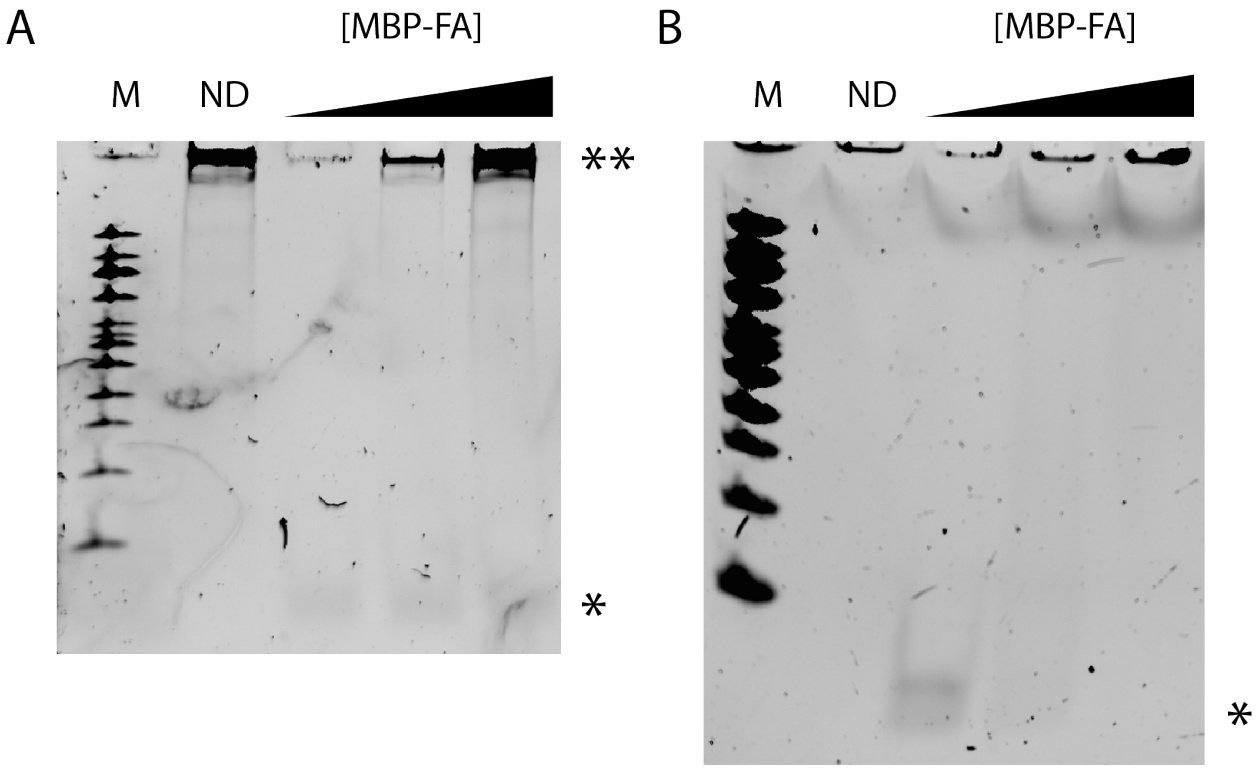

Figure 3.8: DNA binding to purified ComFA - (A)EMSA of MBP-ComFA prior to PEI precipitation. ** Migration location of MBP-ComFA/DNA complex. (B)EMSA of MBP-ComFA following PEI precipitation. (A and B) M: 100 bp ladder, ND: No DNA added. Remaining lanes have increasing amounts of MBP-FA. *Migration location of free ssDNA

protein from C-terminal truncations, which do not precipitate. The resulting protein is apparently DNA-free (Figure 3.8), but is less stable in solution, and appears to aggregate in the absence of a DNA substrate. This aggregation appears to be partially alleviated with the use of small DNA substrates. The protein purified using the PEI precipitation was used to determine the stoichiometries of the wildtype and ComFA^{4CS} fusions.

3.3.6 ComFA binds zinc

After confirming that the putative zinc finger is important for function by *in vivo* transformation efficiency analysis I set out to determine whether ComFA does in-fact bind zinc. To do so I made translational MBP-ComFA fusion constructs for heterologous expression in *E. coli*. I also tested the MBP-ComFA construct in *B. subtilis* for functionality during transformation. The transformation efficiency experiments showed that the fusion does not interfere with ComFA activity (Figure 3.9).

Confirming the zinc-binding activity was performed by a couple methods. First, I had some evidence that metal binding activity was possible from the cross-reactivity observed in the FlAsH experiments (See Appendix B). While examining methods to further purify the MBP-ComFA constructs expressed in *E. coli* I tried immobilizing the MBP-ComFA on an immobilized metal affinity chromatography (IMAC) column charged with Zn^{2+} following cleavage with PreScission Protease. PreScission Protease is a derivative of the 3C human rhinovirus protease that has been translationally fused to GST (127). The protease cleaves a

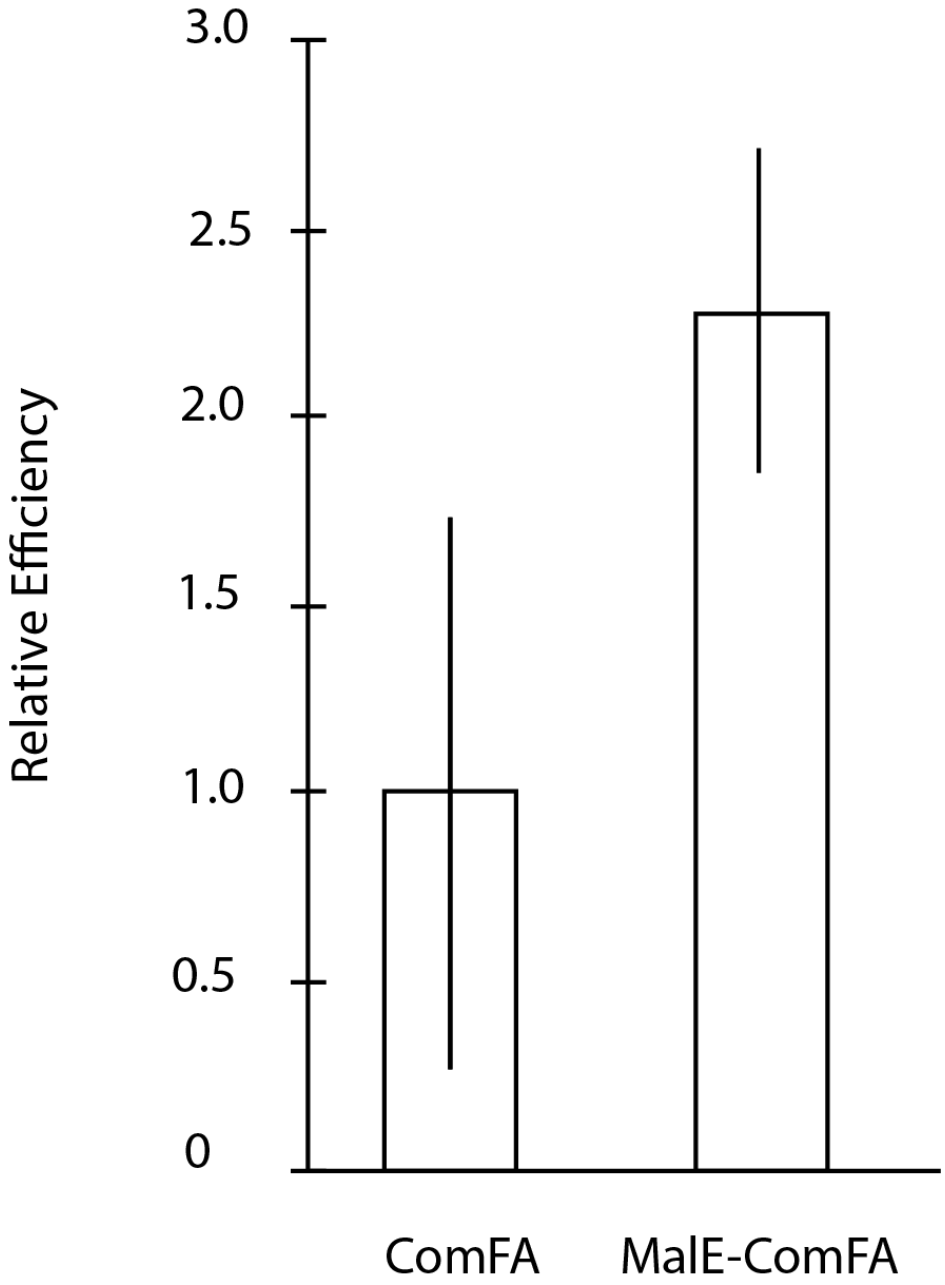

Figure 3.9: MBP-ComFA complements *comF* phenotype - Transformation efficiency for *malE-comFA* fusion relative to wildtype *comFA*. Both constructs expressed from *yvbJ* ectopic locus under the control of a P_{comF} promoter. The *comF* locus has been replaced with a *cat* cassette. Error bars are standard error, wildtype n = 2, *malE-comFA* n = 4.

specific amino acid sequence (LEVLFQ$^{\triangledown}$GP) (127) which I had inserted between the MBP and the ComFA in the MBP-ComFA translational fusion. Passing the cleaved protein over the zinc-IMAC column showed that the cleaved ComFA, and uncleaved MBP-ComFA were retained on the column, and eluted with 250 mM imidazole, while the free and cleaved MBP flowed through the column, and did not bind (Figure 3.11A).

3.3.7 Tetracysteine motif required for zinc-binding activity

I attempted a similar Zn-IMAC experiment using MBP-ComFA^{4CS} to determine if the tetracysteine motif is required for binding to the Zn-IMAC column. I found that even though the PreScission Protease cleavage step failed for some reason, the fusion protein was still mostly present in the flowthrough of the Zn-IMAC column (Figure 3.11B). As the experiment with MBP-ComFA showed that MBP does not mediate interactions with the column the mutation of the cysteines to serine disrupts the interaction of MBP-ComFA^{4CS} with the column (Figure 3.11).

Much later I managed to purify the protein away from the DNA bound to it. Having the DNA-free protein allowed me to test requirement for the tetracysteine motif for binding of zinc ions. I constructed an MBP-ComFA$^{C60S/C67S/C84S/C87S}$ fusion and purified the mutant and wildtype fusion proteins from *E. coli* and tested for zinc content via colorimetric 4-(2-pyridylazo)resoricinol (PAR) assay. I found that the wildtype fusion has a [Zn^{2+}]:[protein] ratio of 0.92±0.04. While the ComFA$^{C60S/C67S/C84S/C87S}$ (ComFA^{4CS}) mutant MBP fusion

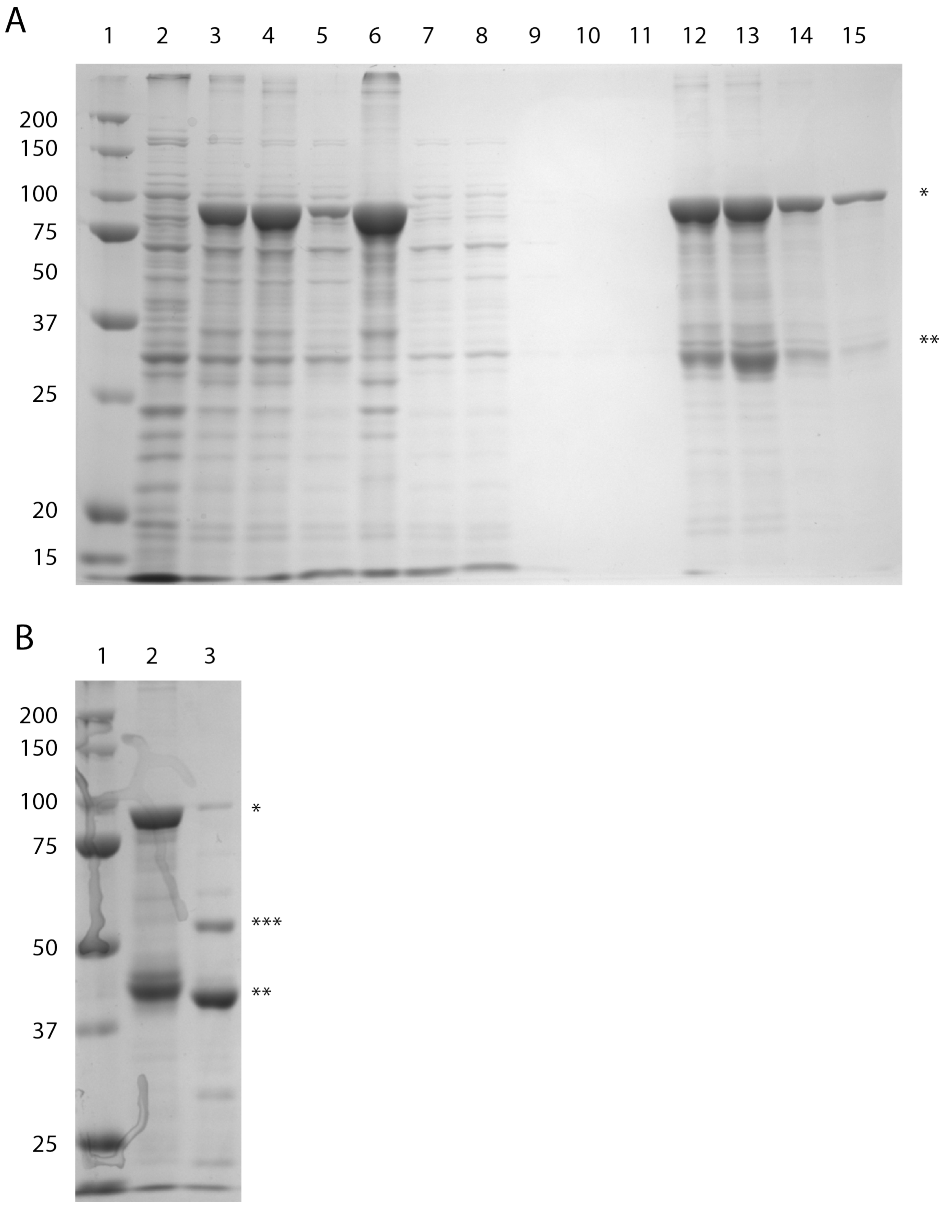

Figure 3.10: Purification of MBP-ComFA for Zn-IMAC - SDS-PAGE gel of steps in purification of ComFA. (A and B)*MBP-ComFA, **MBP. (A) Lysis and sepharose dextrin affinity chromatography (SDAC) of MBP-ComFA. 1: Molecular weight marker; 2: Pre-induction; 3: Post-induction; 4: Lysate; 5: Supernatant following 100 000 x *g* centrifugation; 6: Pellet following 100 000 x *g* centrifugation; 7-8: Flowthrough from SDAC; 9-11: Wash fractions from SDAC; 12-15: Elution fractions from SDAC.(B)Cleavage with Precission Protease. Lane 1: Molecular weight marker; 2: Pooled sepharose dextrin eluate; 3: Eluate following 2-hour Precission Protease cleavage, ***ComFA.

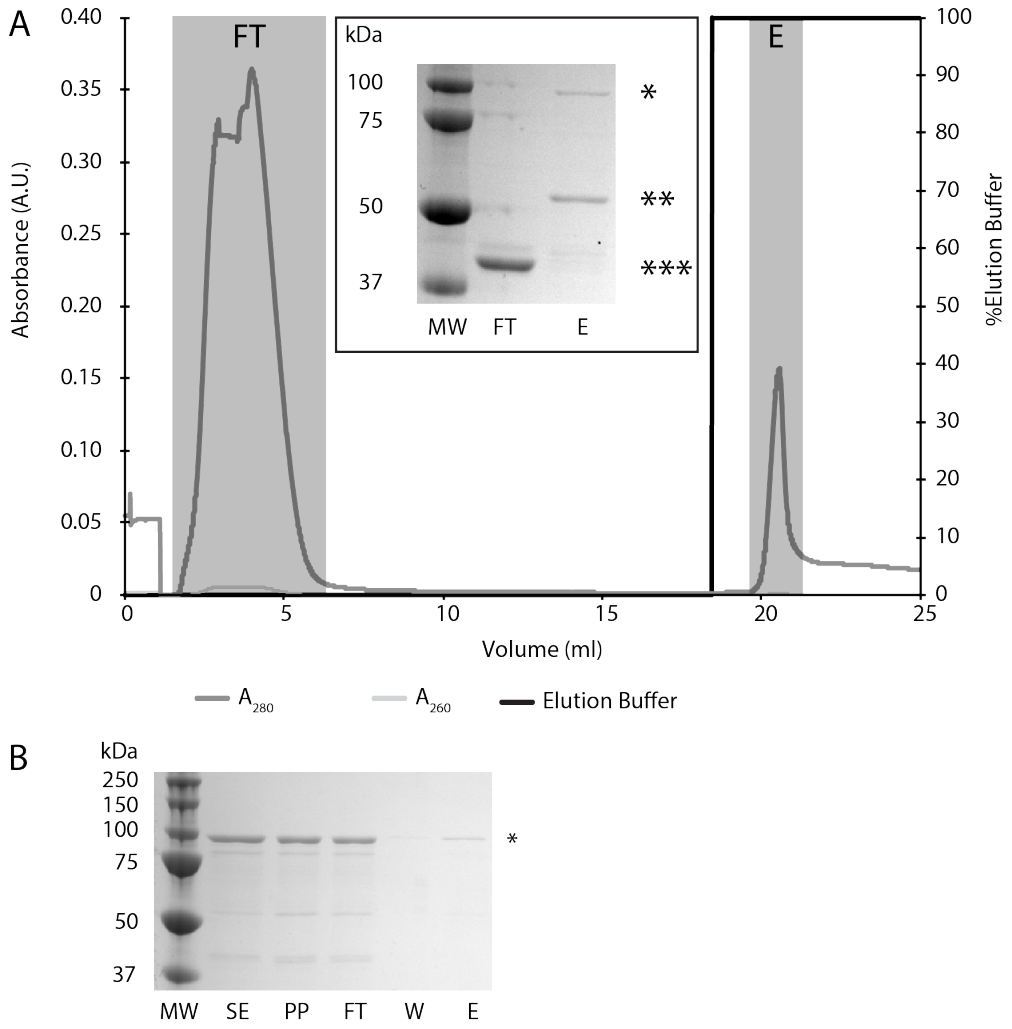

Figure 3.11: Tetracysteine motif is required for binding Zn-IMAC column - (A)Absorbance traces from MBP-ComFA Zn-IMAC. Left axis marks the A_{280} and A_{260} in arbitrary units. The right axis marks the percentage of elution buffer containing 10 mM D-(+)-maltose flowing over the column. The gray boxes show the fractions of the traces that correspond with lanes in inset. Inset: SDS-PAGE gel from MBP-ComFA Zn-IMAC. *MBP-ComFA, **ComFA, ***MBP. MW: Molecular weight marker, FT: Flowthrough, E: Elution (B)SDS-PAGE from MBP-ComFA^{4CS} Zn-IMAC. *MBP-ComFA^{4CS}. MW: Molecular weight marker, SE: SDAC elution, PP: Post-PreScission Protease cleavage, FT: Flowthrough, W: Wash, E: Elution.

had a [Zn^{2+}]:[protein] ratio of 0.44±0.03 (Figure 3.2).

3.3.8 On-going work

The results from the PAR analysis are preliminary, and the assay is still being optimized. Also, additional work is required to demonstrate that the loss of the tetracysteine motif does not modulate the stability of the ComFA mutants *in vivo*. While the phenotype being the result of decreased protein levels is unlikely, a Western blot analysis is required to confirm that the expression levels of the wildtype and mutant alleles are comparable in *B. subtilis*.

3.4 Conclusions

In the work presented here, I have shown that in addition to the DEAD-box helicase motifs ComFA requires zinc-binding for proper function. The requirement for zinc is apparently independent from the DEAD-box helicase activity, as mutants lacking the zinc binding site and carrying mutations in DEAD-box helicase motifs have apparently additive defects.

How this zinc binding site is involved in ComFA activity is still unclear. Zinc fingers have been shown to be involved in a number of processes involving substrate and binding partner recognition, catalysis, and providing structure to protein domains (reviewed in (128)). There are other SF 2 helicases which contain zinc fingers, however, they are generally found in separately folded domains. Given the size of ComFA, and the proximity of the C4-zinc

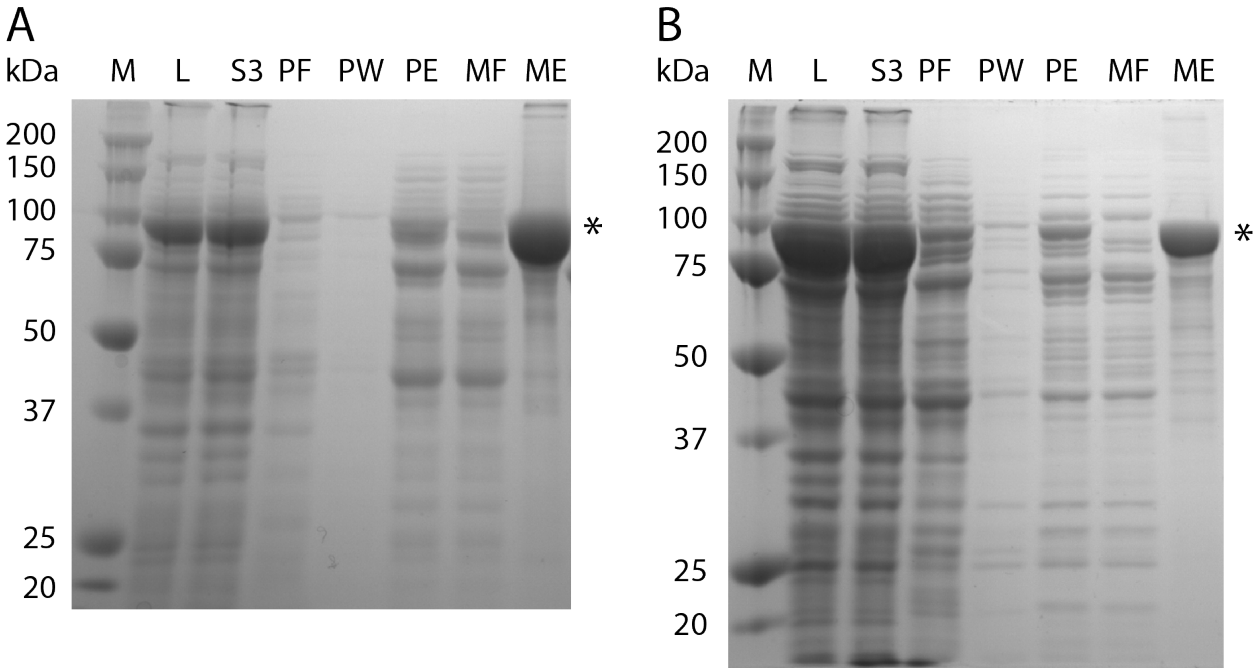

Figure 3.12: MBP-ComFA and MBP-ComFA^{4CS} purifications - Example SDS-PAGE gels for MBP-ComFA and MBP-ComFA^{4CS}(A and B) M: Molecular weight marker; L: Cell lysate; S3: Supernatant from 3 000 x g centrifugation; PF: Flowthrough from PEI precipitation; PW: Supernatant from buffer wash of PEI pellet; PE: Salt elution from PEI pellet; MF: Flowthrough from PE during SDAC; ME: Maltose elution from SDAC. Molecular weight values for bands in in molecular weight marker are indicated to the left of each gel. * Migration location of MBP-ComFA/MBP-ComFA^{4CS} fusion. (A)MBP-ComFA purification. (B)MBP-ComFA^{4CS} purification

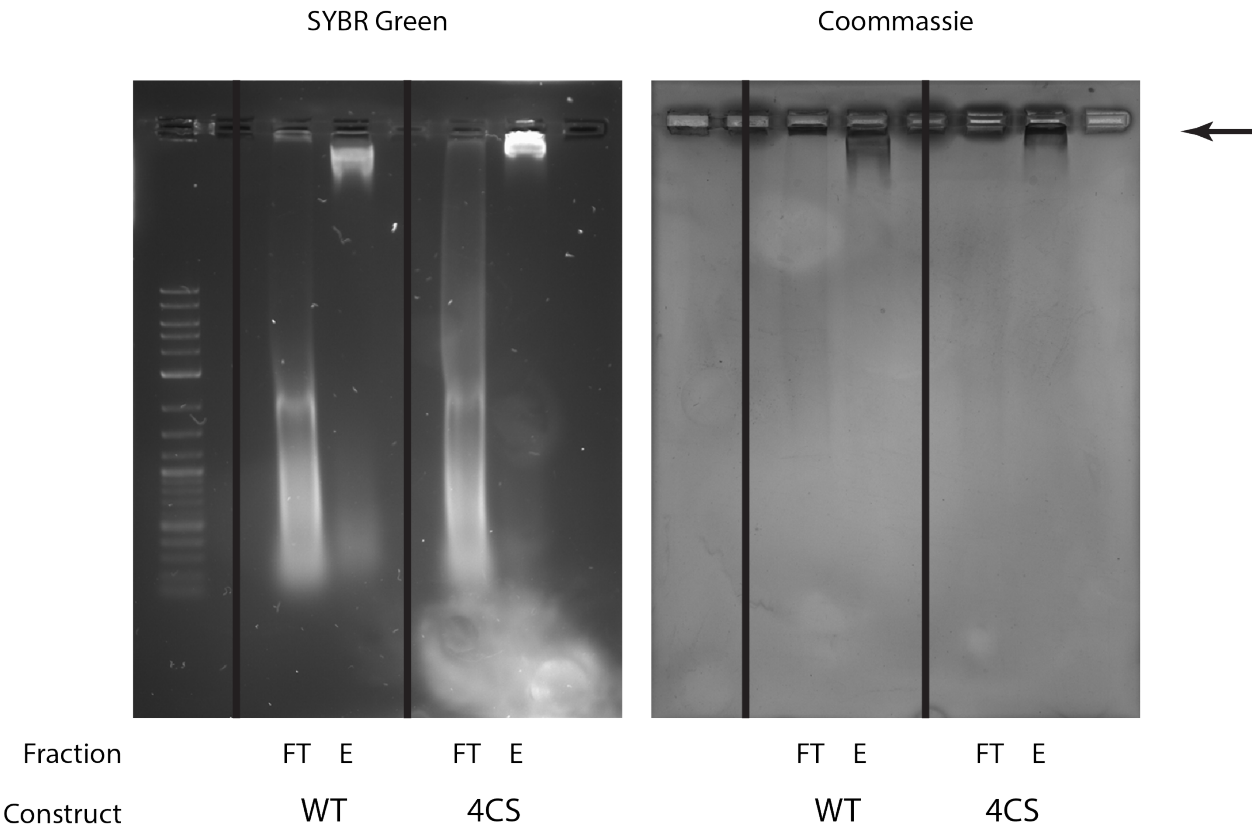

Figure 3.13: MBP-ComFA and MBP-ComFA^{4CS} DNA-binding - 0.8 % Agarose gel with samples of purified MBP-ComFA and MBP-ComFA^{4CS}. FT: flow through from MBPTrap column. E: Elution from MBPTrap column. The arrow denotes the protein-DNA complex. The same gel is depicted twice. On the left, image following staining with SYBR Green DNA stain. On the right, image following Coommassie staining. WT refers to MBP-ComFA and 4CS refers to MBP-ComFA^{4CS}

finger to motif I, it is unclear whether the C4-zinc finger motif is part of a domain that folds independently from the helicase core. C4 zinc fingers have not been shown to be involved in catalytic processes (reviews in (128)), so that leaves intermolecular interactions and intramolecular structure. A number of zinc-containing proteins have been shown to bind DNA, with the zinc finger itself being important for the recognition of a specific sequence (107). *B. subtilis* has shown no preference in the origin of the DNA used in uptake experiments (31, 129). The lack of preference, however, is not a sufficient disqualification as helix-turn-helix motifs like the one found in ComEA and required for DNA-binding are often implicated in sequence-specific interactions as well (54, 55). Furthermore, the involvement of the zinc finger in DNA-binding in ComFA may also be difficult to ascertain as DEAD-box helicases coordinate interactions with DNA substrates through a number of residues along their sequence. That said however, it does not appear that the zinc finger plays a significant role in DNA binding, as we do not observe a difference in DNA bound to purified MBP-ComFA or MBP-ComFA^{4CS} when the PEI precipitation is omitted (Figure 3.13).

Another possibility is that the zinc finger may be involved in protein-protein interactions. This may be more promising given the resilience of DNA-binding in the MBP-ComFA^{4CS}, and the presence of a similar tetracysteine motif in ComFC (58). Examining the protein-protein interactions *in vivo* or *in vitro* may provide some insight into this putative function. Unfortunately, current microscopy techniques performed in *B. subtilis* do not appear to provide sufficient resolution to determine direct interactions (50, 130, 131), and the low

occurrence of competent cells, even in competence growth media, and the fact that several potential interacting partners are membrane proteins, and in some cases expressed at very low copy numbers can make it difficult to determine interactors by *in vitro* methods such as co-immunoprecipitations.

The presence of this zinc binding motif in ComFA and a similar putative motif in ComFC provide some insight into the influence of *znuABC* and *zosA* zinc transporters in the development of competence. An attractive model given these findings is that zinc homeostasis is important for the proper function of these proteins and thus, the zinc transporters are required. To fully develop this model, the other *com* proteins would need to be analyzed for putative zinc binding motifs, and requirement of zinc for function. Also, on a technical note, the competence media used in these experiments calls for additional Mg^{2+} to allow for proper development of competence, and increasing the $[Mg^{2+}]$ can actually improve transformation efficiency. However, any zinc that is likely present in the media is provided by impurities in other components of the media. Since the *in vitro* expression was performed in the presence of 200 μM Zn^{2+} to improve solubility of MBP-ComFA it would be interesting to examine how Zn^{2+} supplementation or depletion of the competence media influences transformation rates.

3.5 Materials and methods

3.5.1 Strains and growth conditions

All *B. subtilis strains* were derived from the prototrophic strain PY79 (92). *B. subtilis* were grown in Luria-Bertani (LB) broth or on LB plates fortified with 1.5 % Bacto agar at 24 °C or 37 °C as appropriate. 10x modified competence (MC) medium was made as described in (93). Cells were grown to competence in 1x MC supplemented with 0.3 % 1 M $MgSO_4$. When appropriate, antibiotics were included at the following concentrations: 5 μg/ml chloramphenicol (Cm_5), 100 μg/ml spectinomycin ($Spec_{100}$), and 1 μg/ml erythromycin plus 25 μg/ml lincomycin (*mls*).

E. coli DH5αwas used to propogate plasmid constructs used in this work. *E. coli* BL21 (DE3) and *E. coli* UT5600 strains were used for protein expression. All *E. coli* strains were grown in LB broth or LB plates fortified with 1.5 % Bacto agar at the indicated temperatures. When appropriate, antibiotics were included in the following concentrations: 50 μg/ml kanamycin (Kan_{50}), 100 μg/ml ampicilin (Amp_{100}).

3.5.2 Plasmid construction

Plasmids used in this work are listed in Table 3.2. Oligonucleotides used in this work are listed in Table 3.3

pSC017 [*malE-comFA*] was generated by a two-way ligation between a *Nde*I-*Bam*HI PCR product containing the *comFA* coding sequence amplified from pBB031 using oSC044 and oSC061 into pMAL-c5E cut with *Nde*I and *Bam*HI.

pSC042 [*malE-3crs*] was generated by a two-way ligation between a *Sac*I-*Hind*III PCR product containing the multiple cloning site of pMAL-c5E and replacing the enterokinase cleave recognition sequence with a 3C human rhinovirus protease cleavage recognition site (3crs) using oSC089, oSC090, oSC091, and oSC095, amplified from pMAL-c5E. The PCR product is inserted into pMAL-c5E cut with *Sac*I and *Hind*III.

pSC045 [*malE-3crs-comFA*] was generated by a two-way ligation between an *Nde*I-*Bam*HI fragment containing the *comFA* coding sequence from pSC017 into pSC042 cut with *Nde*I and *Bam*HI.

pSC048 [P_{comF}-*comFA*] See Chapter 2 Plasmid construction.

pSC097 [h_6-$comFA^{1\text{-}251}$] was generated by a two-way ligation between an *Nde*I-*Bam*HI PCR product containing $comFA^{1\text{-}251}$ amplified from *B. subtilis* genomic DNA using oSC163 and oSC164 into pET28b(+) cut with *Nde*I and *Bam*HI.

pSC098 [h_6-*comFA*$^{1\text{-}252}$] was generated by a two-way ligation between an *Nde*I-*Bam*HI PCR product containing *comFA*$^{1\text{-}252}$ amplified from *B. subtilis* genomic DNA using oSC163 and oSC165 into pET28b(+) cut with *Nde*I and *Bam*HI.

pSC104 [*comF::cat*] See Chapter 2 Plasmid construction.

pSC088 [h_6-*malE-3crs*] was generated by site-directed mutagenesis of pSC042 using oSC138 and oSC139 to insert at 6x His coding sequence upstream of the *malE* coding sequence.

pSC106 [h_6-*malE-comFA*$^{1\text{-}251}$] was generated by a two-way ligation between an *Nde*I-*Bam*HI fragment contain *comFA*$^{1\text{-}251}$ from pSC097 into pSC088 cut with *Nde*I and *Bam*HI

pSC109 [h_6-*malE-comFA*$^{1\text{-}252}$] was generated by a two-way ligation between an *Nde*I-*Bam*HI fragment contain *comFA*$^{1\text{-}252}$ from pSC098 into pSC088 cut with *Nde*I and *Bam*HI

pSC118 [P_{comF}-*comFA*^{C60S}] was generated by site-directed mutagenesis of pSC048 using oSC212.

pSC119 [P_{comF}-*comFA*^{C63S}] was generated by site-directed mutagenesis of pSC048 using

oSC213.

pSC120 [P_{comF}-$comFA^{C84S}$] was generated by site-directed mutagenesis of pSC048 using oSC214.

pSC121 [P_{comF}-$comFA^{C87S}$] was generated by site-directed mutagenesis of pSC048 using oSC215.

pSC129 [$yvbJ::P_{comF}$-$comFA^{C60S}$] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C60S}$ from pSC118 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC130 [$yvbJ::P_{comF}$-$comFA^{C63S}$] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C63S}$ from pSC119 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC131 [$yvbJ::P_{comF}$-$comFA^{C87S}$] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C87S}$ from pSC121 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC137 [$yvbJ::P_{comF}$-$comFA^{C84S}$] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C84S}$ from pSC120 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC219 [P_{comF}-$comFA^{C60S}$ in pMiniMAD2] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C60S}$ from pSC118 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC220 [P_{comF}-$comFA^{C63S}$ in pMiniMAD2] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C63S}$ from pSC119 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC231 [P_{comF}-$comFA^{C84S}$ in pMiniMAD2] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C84S}$ from pSC120 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC232 [P_{comF}-$comFA^{C87S}$ in pMiniMAD2] was generated by a two-way ligation of an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{C87S}$ from pSC121 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC236 [$comFA^{1\text{-}259}$ in pMiniMAD2] was generated by ITA of a PCR product containing $comFA^{1\text{-}259}$ and 424 bases immediately upstream amplified from *B. subtilis* genomic DNA using oSC313 and oSC314, and pMiniMAD2 cut with *Sma*I.

pSC237 [*comFA*$^{218\text{-}463}$ in pMiniMAD2] See Chapter 2 Plasmid construction.

pSC239 [*comFA*$^{218\text{-}463,R419K}$ in pMiniMAD2] See Chapter 2 Plasmid construction.

pSC240 [*comFA*$^{\Delta S1}$ in pMiniMAD2] See Chapter 2 Plasmid construction.

pSC242 [*comFA*$^{1\text{-}259,\ C60S}$ in pMiniMAD2] was generated by ITA of a PCR product containing the upstream of *comFA* contained in pSC236 amplified from *B. subtilis* genomic DNA using oSC313 and oSC335, a PCR product containing *comFA*$^{1\text{-}259,C60S}$ amplified from pSC219, and pMiniMAD2 cut with *Sma*I.

pSC243 [*comFA*$^{1\text{-}259,C84S/C87S}$ in pMiniMAD2] was generated by site-directed mutagenesis of pSC236 using oSC339.

pSC244 [*comFA*$^{1\text{-}259,C60S/C63S/C84S/C87S}$ in pMiniMAD2] was generated by site-directed mutagenesis of pSC236 using oSC338 and oSC339.

pSC262 [P_{comF}-*malE-comFA*] was generated by ITA of a PCR product containing *malE-3crs* amplified from pSC042 and pSC048 cut with *Nde*I.

pSC287 [*malE-3crs-comFA*$^{C60S/C63S/C84S/C87S}$] was generated by site-directed mutagenesis of pSC045 using oSC338 and oSC339.

pSC290 [*comFA*$^{101\text{-}367,E234Q}$ in pMiniMAD2] was generated by ITA of a PCR product containing *comFA*$^{101\text{-}367,\ E234Q}$ amplified from pSC069 using oSC426 and oSC427, and pMiniMAD2 cut with *Sma*I.

pSC291 [*comFA*$^{101\text{-}367,T266A}$ in pMiniMAD2] was generated by ITA of a PCR product containing *comFA*$^{101\text{-}367,\ T266A}$ amplified from pSC216 using oSC426 and oSC427, and pMiniMAD2 cut with *Sma*I.

pSC292 [*comFA*$^{101\text{-}367,S264A}$ in pMiniMAD2] was generated by ITA of a PCR product containing *comFA*$^{101\text{-}367,S264A}$ amplified from pSC229 using oSC426 and oSC427, and pMiniMAD2 cut with *Sma*I.

pSC293 [*comFA*$^{C60S/C63S/C84S/C87S/K152E}$ in pMiniMAD2] was generated by ITA of a PCR product containing *comFA*$^{1\text{-}145,\ C60S/C63S/C84S/C87S}$ and 424 bases immediately upstream amplified from pSC244 using oSC313 and oSC423, a PCR product containing *comFA*$^{138\text{-}259,\ K152E}$ amplified from pSC056 using oSC314 and oSC422, and pMiniMAD2 cut with *Sma*I.

3.5.3 Construction of *B. subtilis* strains

3.5.3.1 Haploid *yvbJ* expression strains

Strains were made *comF* by transformation with genomic DNA from a *comF::cat* strain (bSC049). The *comF::cat* strain was made by the same method as the production of meridiploid strains, except pSC104 was used for integration, *B. subtilis* PY79 was used as host strain, and strains selected for chloramphenicol resistance. Marker replacement verified by PCR amplification of locus using oSC085 and oSC086. Genomic DNA was extracted bSC049 and used for creation of haploid strains. Each host strain was grown at 37 °C for 4 hours in 1x MC. One colony per 1 ml of 1x MC was used for initial inoculum. Dilutions of the genomic DNA gSC019 were made at 1:20 and 1:400 into MiliQ H_2O. At 4 hours 2 μl of each dilution were added to 200 μl of 1x MC culture and grown for an addition 2 hours at 37 °C. After the additional incubation the transformation cultures were plated on LB/Cm_5 agar plates and grown overnight at 37 °C. Following the overnight incubation 8 colonies were chosen and streaked out for single colonies on LB agar plates supplemented with chloramphenicol and incubated overnight at 37 °C. Retention of the original integration was verified by antibiotic selection.

3.5.3.2 *comF* expression strains

See Chapter 2 Materials and methods.

3.5.4 Transformation efficiency analysis

3.5.4.1 *yvbJ* expression strains

Strains were streaked out on LB/Cm_5 or LB/mls plates and cells to be tested grown overnight at 37 °C. *B. subtilis* PY79 control was streaked out on LB as it lacked antibiotic resistance.

3.5.4.2 *comF* expression strains

See Chapter 2 Materials and methods.

3.5.4.3 Genomic DNA preparation and PCR amplification

See Chapter 2 Materials and methods.

3.5.4.4 Colony PCR

See Chapter 2 Materials and methods.

3.5.4.5 Sequencing to verify mutants

Mutations were verified by sequencing of PCR products generated from genomic DNA or colony PCRs.

3.5.5 Statistical Analysis

See Chapter 2 Materials and methods.

3.5.6 Purification of PreScission Protease

Plasmid containing protease and purification protocol received from W. Weihofen (Gaudet Lab). The plasmid (pSC086) was transformed into *E. coli* UT5600. Cells were grown in LB broth supplemented with ampicilin at 32 °C to OD_{600} = 0.3—0.4. The temperature was dropped to 24 °C. Cells were induced with 0.3 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression performed overnight at 24 °C. Cells were harvested by centrifugation at 5 000 rpm in Sorvall SLC-4000 rotor. Cells were resuspended in buffer P1 (50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA) and flash frozen in liquid nitrogen and stored at -80 °C until purification. Cells were lysed in Cell Disruptor (Constant Systems) at 20 kpsi. Lysates were clarified by centrifugation at 100 000 x g for 60 minutes at 4 °C. Lysates were passed over GSTPrep 16/10 column (GE Lifesciences). Column washed with buffer P2 (50 mM Tris-HCl pH 7.5, 150 mM NaCl) and then eluted with buffer P3 (100 mM Tris-HCl pH 8.5, 150 mM NaCl, 15 mM reduced glutathione). Eluate subjected to size exclusion chromatography on an S200 column (GE Lifesciences) equilibrated with buffer P4 (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM EDTA, 5 mM DTT, 20 % glycerol). The protein fractions corresponding with a peak at approximately 70 kDa were collected. The fractions were concentrated to 6 mg/ml and glycerol added to a final concentration of 40 %, and a final concentration of 3 mg/ml. Protein was flash frozen in liquid nitrogen and stored at -80 °C. Concentrated protein was thawed and stored at -20 °C before use.

3.5.7 Expression of MBP-ComFA

3.5.7.1 Enterokinase construct

Expression vector (pSC017) transformed into *E. coli* BL21 cells. Cells were grown in LB broth supplemented with ampicilin, 10 % glycerol and 200 μM $ZnSO_4$ to OD_{600} ~0.2—0.4 at 37 °C then temperature dropped to 16 °C until OD_{600} is ~0.4. Induction with 1 mM IPTG at 16 °C for 16 hours. Cells harvested by pelleting at 5 000 x *g* for 10 minutes, supernatant discarded. Cells resuspended in 20 ml buffer E1 (20 mM Tris-HCl, 200 mM NaCl, 5 % glycerol, 5 mM β-mercaptoethanol, 1 mM Phenylmethanesulfonyl fluoride (PMSF), pH 8 at 4 °C). Cell suspensions dripped into liquid nitrogen and stored at -80 °C until used for purification.

3.5.7.2 Expression for zinc-IMAC analysis

Expression vector (pSC042) transformed into *E. coli* BL21 cells. LB broth supplemented with ampicilin, 10 % glycerol and 200 μM $ZnSO_4$ to OD_{600} ~0.2—0.4 at 37 °C then temperature dropped to 16 °C until OD_{600} is ~0.4. Induction with 1 mM IPTG at 16 °C for 16 hours. Cells harvested by pelleting at 5 000 x *g* for 10 minutes, supernatant discarded. Cells resuspended in 20 ml buffer M1 (50 mM HEPES pH 8, 200 mM NaCl, 5 % glycerol, 5 mM β-mercaptoethanol, 1 mM PMSF). Cell suspensions dripped into liquid nitrogen and stored at -80 °C until used for purification.

3.5.7.3 Expression for PAR analysis

E. coli strain from zinc-IMAC work (eSC052) used. LB/Amp_{100} with 10 % glycerol and 200 μM $ZnSO_4$. Cells grown to OD_{600} ~0.2—0.4 at 37 °C then temperature dropped to 16 °C until OD_{600} is ~0.4. Induction with 1 mM IPTG at 16 °C for 16 hours. Cells harvested by pelleting at 5 000 x *g* for 10 minutes, supernatant discarded. Cells resuspended in 20 ml buffer Z1 (50 mM HEPES pH 8, 200 mM NaCl, 5 % glycerol, 5 mM β-mercaptoethanol,1 mM PMSF). Cell suspensions dripped into liquid nitrogen and stored at -80 °C until used for purification.

3.5.8 Purification of ComFA

3.5.8.1 Enterokinase construct

Frozen cells were thawed on ice. Diluted 1:5 in cold buffer E2 (20 mM Tris-HCl, 200 mM NaCl, 5 mM β-mercaptoethanol, 5 % glycerol). Cells were lysed in One Shot Cell Disruptor (Constant Systems) at 20 kpsi, two passes. Lysate spun at 100 000 x *g* for 1 hour at 4 °C. Clarified cell lysate was passed over sepharose-dextrin resin (GE Lifesciences), equilibrated with buffer E2. Column was washed with buffer M1. Protein was eluted with buffer E3 (20 mM Tris-HCl, 200 mM Nacl, 5 mM β-mercaptoethanol, 5 % glycerol, 10 mM D-(+)-maltose) Elutions were pooled, flash frozen in liquid nitrogen, and stored at -80 °C.

3.5.8.2 For zinc-IMAC analysis

Frozen cells were thawed on ice. Diluted 1:5 in cold buffer M2 (50 mM HEPES pH 8, 200 mM NaCl, 5 mM β-mercaptoethanol, 5 % glycerol). Cells were lysed in One Shot Cell Disruptor (Constant Systems) at 20 kpsi, two passes. Lysate spun at 100 000 x *g* for 1 hour at 4 °C. Clarified cell lysate was passed over sepharose-dextrin resin (GE Lifesciences), equilibrated with Buffer M2. Column was washed with buffer M2. Protein was eluted with buffer M3 (50 mM HEPES pH 8, 200 mM Nacl, 5 mM β-mercaptoethanol, 5 % glycerol, 10 mM D-(+)-maltose) Elutions were pooled, flash frozen in liquid nitrogen, and stored at -80 °C.

3.5.8.3 For PAR analysis

For purification, cells thawed on ice. All buffers treated with Chelex-100 (Bio-Rad Laboratories). Diluted 5-fold into chilled and Chelex-100 (Bio-Rad Laboratories) treated buffer A. Diluted 1:5 in cold buffer Z2 (50 mM HEPES pH 8, 200 mM NaCl, 5 mM β-mercaptoethanol, 10 % glycerol). Cells were lysed in One Shot Cell Disruptor (Constant Systems) at 12.7 kpsi, two passes. Lysate was spun at 3 000 x *g* for 10 minutes at 4 °C to pellet unbroken cells and debris. Protein and DNA precipitated polyethylenimine (PEI) precipitation method modified from (132). PEI added to 0.12 % from a 6 % (v/v) stock and incubated on ice for 15 minutes. Precipitate was pelleted at 3 000 x *g*. Pellet was washed with an equal volume of buffer Z2. MBP-ComFA was eluted by resuspendind the pellet in buffer Z3 (50

mM HEPES pH 8, 500 mM NaCl, 5 mM β-mercaptoethanol, 10 % glycerol) and centrifuging at 10 000 x g for 10 minutes at 4 °C. PEI Eluate was passed over a sepharose dextrin column and washed with buffer Z3. MBP-ComFA was eluted from sepharose dextrin using buffer Z4 (50 mM HEPES pH 8, 500 mM NaCl, 5 mM β-mercaptoethanol, 10 % glycerol, 10 mM D-(+)-maltose). Eluate from sepharose dextrin column was pooled and concentrated before use in PAR analysis.

3.5.9 Enterokinase proteolysis

Enterkinase purchased from New England Biolabs as 2 μg/ml solution. Thawed MBP-ComFA incubated with (w/w) ratios of Enterokinase. Tested as 5 μg aliquots of MBP-ComFA. Incubated at 24 °C for 1—3 hours. Supernatant and pellet fractions were tested following centrifugation at 20 817 x g for 10 minutes at 4 °C. Samples taken and resolved by SDS-PAGE and stained using Bio-Safe Coomassie (Bio-Rad Laboratories, Inc).

3.5.10 PreScission Protease proteolysis

Protein content of eluate was estimated by A_{280} and MBP-ComFA extinction coefficient. 3 mg/ml purified PreScission Protease was added in 1:100 w/w and incubated at RT for 2 hours.

3.5.11 Zinc-IMAC analysis

Following proteolysis sample loaded onto IMAC FF column (GE Life Sciences) charged with $ZnSO_4$. Chromatography performed as in (133) Column was equilibrated with buffer M2, and washed with same buffer. Protein eluted with buffer M4 (50 mM HEPES pH 8, 200 mM Nacl, 5 mM β-mercaptoethanol, 5 % glycerol, 10 mM D-(+)-maltose, 250 mM imidazole-HCl pH 8).

3.5.12 EMSA analysis

Purified MBP-ComFA was incubated with 50 nM oSC143 for 1 hour on ice. Reaction ended with 1:10 addition of 10x loading dye. Run in 5 % 37.5:1 Acrylamide/Bis- for 2 hours in 0.1x TAE buffer at 4 °C. Stained with 1: 10 000 SYBR Green dye for 10 minutes and imaged using trans-UV illumination.

3.5.13 PAR colormetric analysis

Zinc content analysis performed as previously described in (101). Protein concentration determined by unfolding protein by dilution 1:5 in 6 M guanidine-HCl and measuring the A_{280} and using the theoretical extinction coefficient for the fusion protein constructs. Protein mixed 1:4 with 6 M chelex-treated guanidine-HCl to facilitate release of Zn^{2+} prior to PAR addition. The extinction coefficient for $(PAR)_2 \bullet Zn^{2+}$ calculated from a standard curve using

$ZnSO_4$. Zn^{2+} concentration determined from $(PAR)_2 \bullet Zn^{2+}$ extinction coefficient.

Table 3.1: Strains used in Chapter 3

Strain	Genotype	Reference
B. subtilis PY79		(92)
bSC007	*comFA::Tn524 (erm)*	This work
bSC049	*comF::cat*	This work
bSC140	*comF::cat*; *yvbJ::*P_{comF}*-comFA (erm)*	This work
bSC166	*comF::cat*; *yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$ *(erm)*	This work
bSC168	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C60S} *(erm)*	This work
bSC170	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C63S} *(erm)*	This work
bSC172	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C84S} *(erm)*	This work
bSC174	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C87S} *(erm)*	This work
bSC185	*comFA*^{C63S}	This work
bSC196	*comFA*^{C84S}	This work
bSC197	*comFA*^{C87S}	This work
bSC205	*comFA*^{C60S}	This work
bSC207	*comFA*$^{C84S/C87S}$	This work
bSC209	*comFA*$^{C60S/C63S/C84S/C87S}$	This work
bSC229	*comFA*$^{C60S/C63S/C84S/C87S/R419K}$	This work
bSC292	*comFA*$^{C60S/C63S/C84S/C87S/K152E}$	This work
bSC294	*comFA*$^{C60S/C63S/C84S/C87S/S264A}$	This work
bSC296	*comFA*$^{C60S/C63S/C84S/C87S/E234Q}$	This work
bSC298	*comFA*$^{C60S/C63S/C84S/C87S/T266A}$	This work
E. coli BL21 (DE3)	F$^-$ *ompT gal dcm lon hsdSB(rB*$^-$ *mB*$^-$*) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	
E. coli DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169*, hsdR17(r_K^- m_K^+), λ–	
E. coli UT5600	F- *ara-14 leuB6 secA6 lacY1 proC14 tsx-67 Δ(ompT-fepC)266 entA403 trpE38 rfbD1 rpsL109 xyl-5 mtl-1 thi-1*	
eSC052	*E. coli* BL21 (DE3) pSC045 (*bla*)	This work
eSC097	text it E. coli UT5600 pSC086 (*bla*)	This work
eSC188	*E. coli* BL21 (DE3) pSC109 (*bla*	This work

Continued on next page

Table 3.1 – *Continued from previous page*

Strain	Genotype	Reference
eSC207	*E. coli* BL21 (DE3) pSC287 (*bla*)	This work

Table 3.2: DNA & Plasmids used in Chapter 3

Name	Features	Reference
pBlueScriptSK(+)	*bla*	
pET28b(+)	*kan*	Novagen
pMiniMAD2	*bla erm*	(97)
pBB031	$P_{T7}H_6$-*comFA kan*, derived from pET28b(+)	B.M. Burton
pBB268	*bla erm*	B.M. Burton
pBB278	*yhdGH::spec bla*	B.M. Burton
pMAL-c5E	*malE bla*	New England Biolabs
pSC017	*malE-comFA*, derived from pMALE-c5E	This work
pSC042	*malE-3crs* 3C cleavage site, derived from pMAL-c5E	
pSC045	P_{tac}-*malE-comFA*, derived from pSC042	This work
pSC048	P_{comF}-*comFA* derived from pBlueScriptSK(+)	This work
pSC086	GST-3CHRV Protease *bla*	Gift from R. Gaudet
pSC088	h_6-*malE*, derived from pSC042	This work
pSC097	h_6-$comFA^{1\text{-}251}$, derived from pET28b(+)	This work
pSC098	h_6-$comFA^{1\text{-}252}$, derived from pET28b(+)	This work
pSC104	*comF::cat bla*	This work
pSC106	P_{tac}-h_6-*malE*-$comFA^{1\text{-}251}$, pSC088 with insert from pSC097	
pSC109	P_{tac}-h_6-*malE*-$comFA^{1\text{-}252}$, pSC088 with insert from pSC098	This work
pSC118	$comFA^{C60S}$, derived from pSC048	This work
pSC119	$comFA^{C63S}$, derived from pSC048	This work
pSC120	$comFA^{C84S}$, derived from pSC048	This work
pSC121	$comFA^{C87S}$, derived from pSC048	This work

Continued on next page

Table 3.2 – *Continued from previous page*

Name	Features	Reference
pSC129	*yvbJ::*P_{comF}*-comFA*^{C60S}, derived from pBB268 with insert from pSC118	This work.
pSC130	*yvbJ::*P_{comF}*-comFA*^{C63S}, derived from pBB268 with insert from pSC119	This work
pSC131	*yvbJ::*P_{comF}*-comFA*^{C87S}, derived from pBB268 with insert from pSC121	This work
pSC137	*yvbJ::*P_{comF}*-comFA*^{C84S}, derived from pBB268 with insert from pSC120	This work
pSC219	pMiniMAD2 with insert from pSC118	This work
pSC220	pMiniMAD2 with insert from pSC119	This work
pSC231	pMiniMAD2 with insert from pSC120	This work
pSC232	pMiniMAD2 with insert from pSC121	This work
pSC236	pMiniMAD2 with *comFA*$^{1\text{-}259}$ and 424 bases upstream	This work
pSC237	pMiniMAD2 with *comFA*$^{218\text{-}463}$ and 472 bases downstream	This work
pSC239	*comFA*^{R419K}, derived from pSC237	This work
pSC240	pMiniMAD2 with insert derived from pSC081	This work
pSC242	*comFA*^{C60S}, derived from pSC236	This work
pSC243	*comFA*$^{C84S/C87S}$, derived from pSC236	This work
pSC244	*comFA*$^{C60S/C63S/C84S/C87S}$, derived from pSC236	This work
pSC262	*malE-comFA*, *malE* from pSC042 inserted into pSC048	This work
pSC266	pBB268 with insert from pSC262	This work

Continued on next page

Table 3.2 – *Continued from previous page*

Name	Features	Reference
pSC287	*malE-comFA*$^{C60S/C63S/C84S/C87S}$, derived from pSC045	This work
pSC290	*comFA*$^{101\text{-}367,E234Q}$ pMiniMAD2 with insert from pSC069	This work
pSC291	*comFA*$^{101\text{-}367,T266A}$, pMiniMAD2 with insert from pSC216	This work
pSC292	*comFA*$^{101\text{-}367,S264A}$, pMiniMAD2 with insert from pSC229	This work
pSC293	*comFA*$^{C60S/C63S/C84S/C87S/K152E}$, derived from pSC244	This work
pUC19	*bla*	

Table 3.3: Oligonucleotides used in Chapter 3

Name	Sequence[a,b]
oSC024	CATATTCTTTATGCCGGCCGGTT**T**TTCCT GCAATTTGAACAAGTGCGCT
oSC085	TGTATCCATTTGACTCAGAGATCAGC
oSC086	TCTCTGATCCTTGTTCTCCACACC
oSC089	ACTAATTCGAGCTCGCTGGAAGTTCTGTT CC
oSC090	CCCATGGACATATGGGGCCCCTGGAACAG AACTTCCAGCGAGCT
oSC091	GGCCCCATATGTCCATGGGCGGCCG
oSC095	TGTCCTACTCAGGAGAGCGTTCA
oSC138	GTGATGATGGTGGTGATGGCTGCTGCCC ATAATCTATGGTCCTTGTTGG
oSC139	CCATCACCACCATCATCACAGCAGCGGCA TGAAAATCGAAGAAGGTAAAC
oSC163	GGAATTCCATATGAATGTGCCAGTTGAAA AAAACAGTTCC
oSC164	CGGGATCCTATCATTATTGAACAGCGAAT TGAAGGGTTTGATC
oSC212	CTATATCCATAAATAAACGGAGATATAGA **AGC**AATAGGTGCGGACAAACTGATC

Continued on next page

Table 3.3 – *Continued from previous page*

Name	Sequence
oSC213	CCATAAATAAACGGAGATATAGATGTAAT AGG**A**GCGGACAAACTGATCAGCGG
oSC214	CACTCATCTGGAAAGAATAAGCTGTAT**A**G CCGTTCCTGTGTCATGATGG
oSC215	GGAAAGAATAAGCTGTATTGCCGTTCC **AGC**GTCATGATGGGCAGAGTGAGTG
oSC311	CAGGTCGACTCTAGAGGATCCCCCAGCTT TTGCGATATAAAGATGCAATC
oSC312	GTGAATTCGAGCTCGGTACCCGATTTTCT TTAATTTGCTTCTGCAAGAATAAC
oSC313	CAGGTCGACTCTAGAGGATCCCCCAAGCC TTCATTGGTAGTCTTCTAAAGGTAAAG
oSC314	GTGAATTCGAGCTCGGTACCCGGTGCTGT TTTTCTTTCTTGCTTTTTG
oSC335	CTGTTTTTTTCAACTGGCACATTCACATA GCACGCCTCCTTTCGAAAC
oSC338	CCATAAATAAACGGAGATATAGA**A**G**C**AA TAGG**A**GCGGACAAACTGATCAGC
oSC339	CATCTGGAAAGAATAAGCTGTAT**A**GCCGT TCC**A**G**C**GTCATGATGGGCAGAGTG
oSC400	ACTGTTTCGAAAGGAGGCGTGCCATATGA AAATCGAAGAAGGTAAACTGG
oSC401	GAACTGTTTTTTTCAACTGGCACATTCAT ATGGGGCCCCTGG
oSC426	CAGGTCGACTCTAGAGGATCCCCTCATGG AAAGAGGAAAATGAATCAAAC
oSC427	GTGAATTCGAGCTCGGTACCCCTTTCTAT GCTTGTCTTCCGCGTG

[a]Bold-face indicates mutagenic residues.
[b]Underlines indicate restriction sites.

4

Discussion

4.1 Broader significance

Transformation and competence are interesting biological processes on their own, however studying them can give us greater understanding of larger biological problems. The study of transformation can provide insights into the importance of recombination, including questions in eukaryotes, how cells solve problems relative to non-vesicle macromolecular transport, and provide potential technological advances to enable genetic studies of other organisms.

4.2 The importance of competence

There are three main theories regarding the purpose of competence in bacteria: a method for scavenging DNA as a nutrient source, a method for facilitating DNA repair following

insults to the genome, and a method for allowing genetic resorting, and sampling the fitness landscape.

4.2.1 DNA as food

Part of the argument for DNA being a potential nutrient source is based upon what we know of the atomic make up of biological molecules and DNA. Generally, carbon, hydrogen, nitrogen, oxygen, phosphorous, sulfur, and metals are required to make the molecules which make up the macromolecules that compose living organisms. DNA contains carbon, hydrogen, nitrogen, oxygen, and phosphorous, almost all of those necessary nutrients. While the triggers for the development of competence vary, in *B. subtilis* it develops as part of a starvation response. *Bacillus* species have also been show to make biofilms which contain DNA in the extracellular matrix (134, 135). So the logic goes that it would be greatly advantageous to have a system which allows you to utilize the DNA present in the environment around you, when you are lacking other sources of nutrients.

When we think about competence more generally, the regulatory picture does not seem to reflect that of a starvation response. Different bacterial species have evolved distinct signaling pathways to govern the expression of competence genes (25, 136, 137, 138, 139, 140, 141). In *B. subtilis* only a fraction of the cells, about 10% become competent. When we examine this competent population of cells we find a significant number of physiological changes which even modify the buoyancy of competent cells relative to non-competent cells. These factors

suggest something more complex than simply using DNA as a nutrient source.

Another component of the process that argues against the uptake machinery being a scavenging system is that it is specific to DNA transport (i.e. RNA and synthetic polymers are not transported) (28, 129), and only one strand is imported into the cell while the other is degraded, which seems like a horrible waste of resources (142, 143, 144, 145). That said, the purpose of degrading one strand is still unclear. Without going into too much detail, importing ssDNA is thought to provide some protection to the cell by preventing transcription, protecting the substrate by making it a ssDNA-binding protein substrate, or a preparation for recombination (146).

4.2.2 DNA uptake for repair

One of the challenges bacteria face is that they are haploid. When diploid organisms undergo DNA damage, they have another copy of their genes, which can be used as a template for homologous recombination. One potential use for transformation is thought to be a way of solving the problem. Essentially, if cells are damaged due to an environmental insult some cells would be expected to perish, releasing their genetic material into the environment. Competent cells that sustained damage, but were not killed could use the DNA now in the environment as material to provide a template for homologous recombination.

Furthermore, recombination proteins are upregulated during competence, and there is some evidence that the increase may help cells cope with genetic insults. Which fits well

with the quorum regulation of competence in some species (37, 147), and the self-recognition sequences found in *N. gonorrhoeae*, for example (148). However, it does not appear that the competence machinery is upregulated as part of a DNA damage response.

4.2.3 DNA uptake for genetic exchange

Horizontal gene transfer mechanisms are sometimes referred to as bacterial sex, as they allow genetic exchange in ways otherwise restricted to sexually reproducing organisms. Many of the concepts we are familiar with from eukaryotic genetics such as independent assortment which are based upon recombination manifest in very different ways in bacteria.

Generally bacteria are haploid and hold their genes on a single chromosome, though plasmids are also sometimes found as well. Furthermore they reproduce asexually with a parent cell that divides into two daughter cells. As such there are not regions of homology within the genome for homologous recombination, with a few exceptions.

If we take a moment to think about how evolution functions in eukaryotes, an asexually reproducing bacterial species lacking recombination is equivalent to an extremely inbred population–they have very little genetic diversity. Genetic diversity is introduced into the population is through replication errors and repair errors following DNA damage. Mutations in the genome are detrimental to protein function more frequently than they are beneficial to protein function. Selection in the environment would control the most deleterious mutations as the individuals containing those would have a large disadvantage manifested as reduced

growth rates or increased mortality rates. Even so, overtime we would expect that the number of deleterious mutations in a genome would increase, decreasing the relative fitness of a population relative to its ancestors in this scenario. This phenomenon is known as Mueller's ratchet. If we assume limited competition, any individual mutation which is severely deleterious would of course be quickly removed from the population, however, marginally deleterious mutations could accumulate in the population over time.

Where Muller's ratchet becomes very important is when we think about beneficial mutations. If we build upon the scenario above, in which we have a population which has accumulated a number of mutations overtime such that there is genetic variation among individuals, as well as variance in the frequency of alleles at loci throughout the genome then each individual has a level of fitness determined by the combined influences of the genes it carries. At some rate lower than the rate of mutations from replication, we can observe some individuals undergoing mutations which provide a competitive advantage. In order for those beneficial mutations to manifest in changes in the composition of the population their fitness improvement must outweigh the current mutation burden when they arise.

Even if we ignore the burden complication, the alleles face another challenge to fixing in the population: each other. As any mutation which arises is unlikely to arise in the same locus multiple times we can assume that any beneficial allele would be unique in the population. Thus it follows that these alleles would compete with each other in a population. And in the absence of recombination a cell would need a beneficial mutation to arise in a

genome already carrying another in order to harness the benefits of both. A second beneficial mutation is most likely to occur once another beneficial mutation is a significant fraction of the population, or has taken over (Figure 4.1ii). With that in mind, they are thought to arise sequentially. The process described above is know as clonal interference. The independent competitive advantages provided by each beneficial mutation must battle it out to dominate the population and thus "interfere" with each other, and slow the rate of evolution.

Genetic exchange allows the alleviation of clonal interference. If we take the above population and allow for genetic exchange by individuals by competence then if beneficial alleles "A" and "B" arise independently in a population that initially was ab we can see the presence of AB individuals much sooner than without genetic exchange (Figure 4.1). In the simplest scenario we can have an individual of aB which perishes as the result of either gene a or an unrelated cause and release genetic material containing B. If a competent individual of genotype Ab were to take up the genetic material containing B the resolution of recombination can result in an individual with the genotype AB which would now benefit from the growth advantages of both A and B (Figure 4.1i). Moradigaravand *et al.* performed an interesting set of simulations that nicely illustrate the potential advantages of competence and recombination (11). There are also examples from *in vivo* experiments which suggest that competence provides a number of selective advantages with regard to the maintenance and sharing of beneficial alleles (24).

Recombination and HGT alleviate clonal interference by allowing genes to assort be-

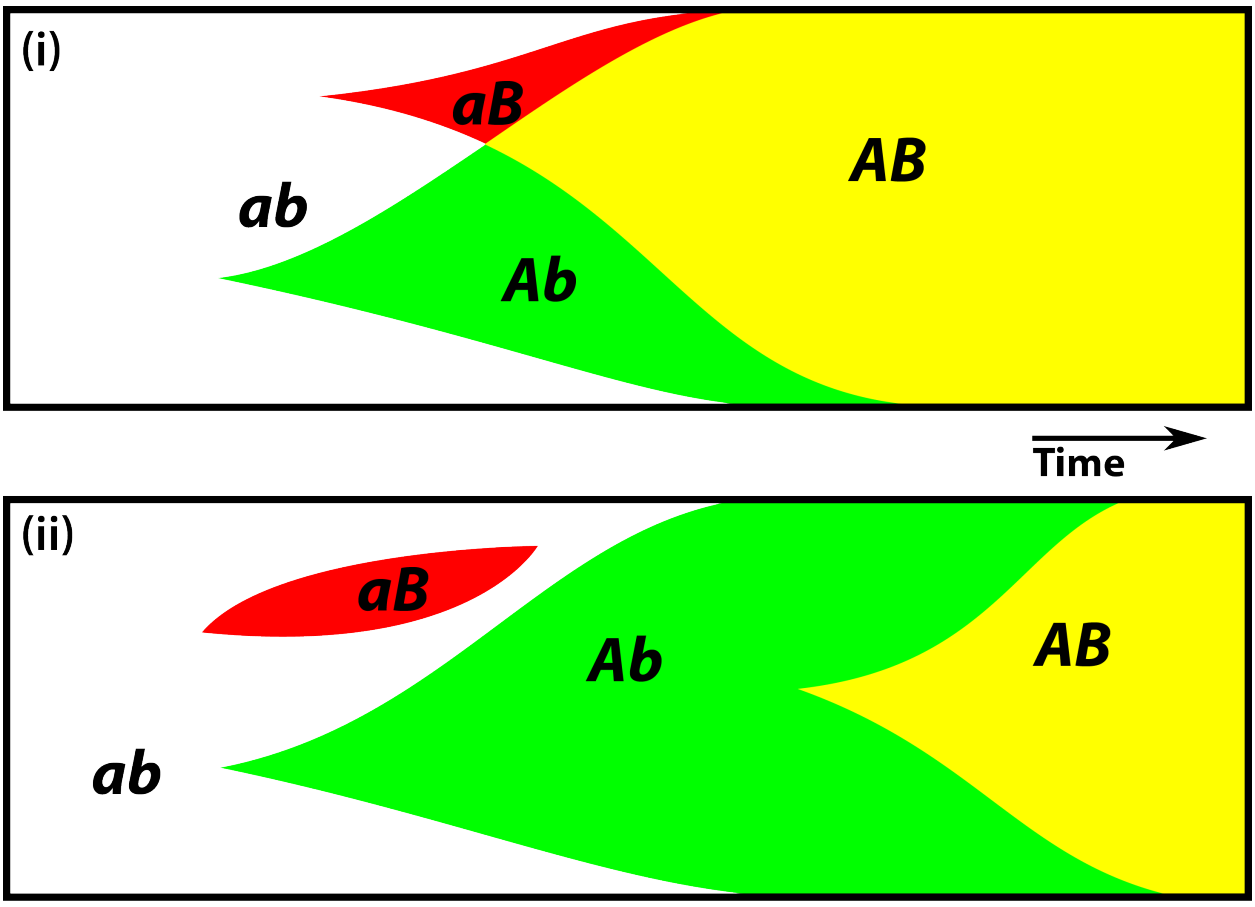

Figure 4.1: Clonal interference - (i and ii) "A" and "B" refer to beneficial alleles at their respective loci. "a" and "b" refer to the ancestral alleles at their respective loci. (i) Allele distribution overtime in a population, allowing for recombination. (ii) Allele distribution overtime in a population where recombination is prohbited. Adapted from:http://en.wikipedia.org/wiki/File:Evolsex-dia2a.svg.

tween cells in a population. Therefore beneficial alleles can be combined and can propagate through the population at a higher frequency. In microbial communities species such as *B. subtilis* that are promiscuous with respect to the sources of DNA they take up also gain the opportunity to obtain new genes, and gene networks that were previously absent from their genomes.

4.3 ComFA and transformation

So you may be asking "What does all this have to do with ComFA?" If you remember that ComFA is present in Gm^+ which lack a periplasm and outer membrane and absent in Gm^-s which have a periplasm and outer membrane, then competence appears to be largely a mechanism for sampling the fitness landscape. Having a cytosolic helicase, even a non-processive one, greatly improves a cell's ability to import high-quality recombination substrate when it lack a protective compartment in the event of asynchronous import steps.

The general model for DNA uptake in *B. subtilis* has assumed a rapid, processive process underlies the observed uptake rates. However, there is evidence that suggests that the internalization process in *B. subtilis*, and possibly other Gm^+ bacteria, occurs in two steps (47, 48), similar to what is observed in Gm^- bacteria (149). The DNase-resistant state during uptake appears to precede the transport across the cell membrane (46, 48), which also confers cyanide resistance (47). In *H. pylori* it appears that DNA transport across the outer membrane is not coupled to transport across the inner membrane, and in the absence

of inner membrane transport DNA can accumulate in the periplasm (149).

It is possible that the function of the pseudopilus, believed to be involved in pulling DNA to the membrane *B. subtilis* (43, 48, 53, 55), may not be coupled to any activity from the cytoplasmic components of the uptake machinery. Thus, it is worth noting that while ComFA is found in a number of Gm^+ bacterial species no orthologs have been identified in Gm^- bacteria. With that in mind, why would the Gm^+ bacteria need a dedicated helicase, while single-stranded DNA-binding proteins and the recombination machinery are sufficient to facilitate cytosolic transport in Gm^- bacteria?

The first thing to consider is the structural differences between the two cell types. Gm^- bacteria contain a second membrane absent in Gm^+ bacteria which creates the periplasm compartment. The periplasm may provide a protected space in which the incoming DNA can accumulate in the event that cytoplasmic transport lags behind import into the periplasm. Without this protected space *B. subtilis* and other Gm^+ bacteria may need a faster transport mechanism to ensure that the integrity of the incoming DNA is maintained. Furthermore Gm^+ bacteria contain a membrane-bound nicking endonuclease (NucA) which is required to create the double-strand breaks to liberate the necessary free-end to allow transport. The nicking activity of NucA appears to be regulated by the presence of substrate. It would be reasonable to assume that the longer the incoming DNA is allowed to reside at the surface of the cell, the more nicks would be introduced. The highly fragmented DNA would be a poor substrate for subsequent recombination following import into the cytoplasm.

4.4 ComFA is a DEAD-box helicase with a metal-binding motif

In the previous two chapters I have detailed my examination of the DEAD-box helicase motifs and a metal-binding motif present in ComFA. The DEAD-box motifs and the metal-binding motif are required for ComFA function, and independently contribute to the function of ComFA (see Figures 2.3, 3.2A, 3.3). With this information we can conclude that ComFA is in fact a DEAD-box helicase, and that it contains an accessory metal-binding motif which is capable of binding zinc, and transformation in Gm^+ bacteria is most likely cytosolically powered by a non-processive helicase. This does not necessarily say that ComFA does not function in a processive manner, but does indicate that it likely is not intrinsically a high-processivity helicase.

The results from the experiments described in Chapters 2 & 3 support the following model. In the wildtype case, both the helicase and zinc finger functions are intact and contribute to wildtype transformation efficiency (Figure 4.2A and see Figure 3.3). The *comFA*$^{\Delta S1}$ allele creates a non-functional ComFA. In this background DNA uptake is likely mediated by the ssDNA binding proteins, and *rec* machinery normally downstream of ComFA (Figure 4.2B and see Figure 3.3). The similarity in transformation efficiency for the zinc finger/helicase combination mutants to *comFA*$^{\Delta S1}$ suggests that these point mutants create

a similar functional phenotype for ComFA (see Figure 3.3). The DEAD-box helicase mutants only have the zinc finger intact (zinc finger only) and therefore cannot effectively translocate along the incoming DNA. They should, however, maintain their ability to bind the DNA and behave as a Brownian ratchet to facilitate DNA import by binding the DNA as it comes into the cell and preventing the DNA from passing back through ComEC once it has entered the cytoplasm (Figure 4.2C). The function of the zinc finger is not clear, however, given that its function is independent of the helicase activity (see Figure 3.3) and it is not required for DNA binding (see Figure 3.13). If its function was related to the helicase activity, we would expect that the helicase mutants would account for the entire loss of activity, and would be similar to *comFA*$^{\Delta S1}$ in the transformation efficiency experiments, not unlike the observations found in (49). Thus, it is reasonable to assume that it is not involved in substrate recognition or enhancement of helicase activity. As such, when the zinc finger is mutated, the ComFA^{4CS} (helicase only) likely maintains its helicase activity, but is unable to perform another function, such as properly localizing to the sites of DNA uptake, or oligomerizing, interacting with other proteins necessary for efficient uptake (Figure 4.2D).

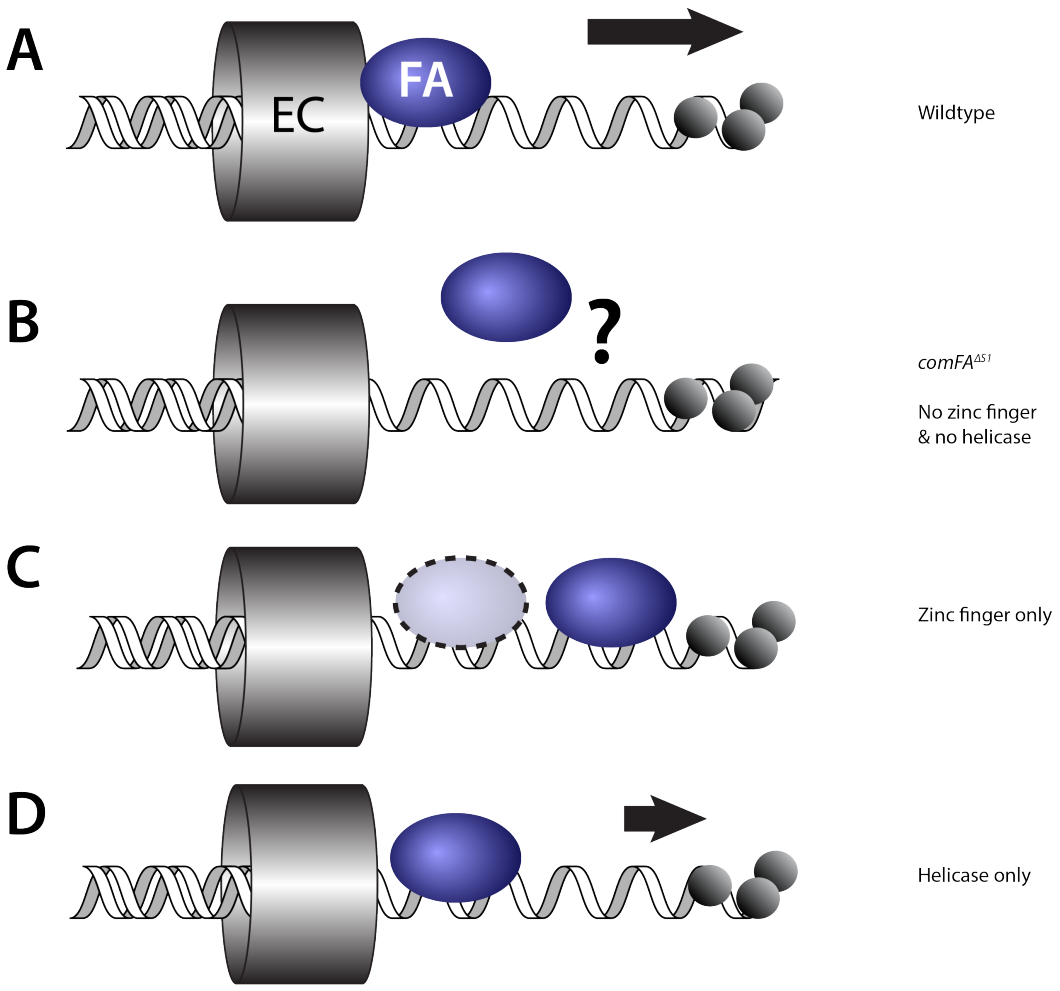

Figure 4.2: Model of ComFA mutant function during transformation - Schematic models of transformation in *comFA* mutant backgrounds. See Figure 3.3 for corresponding transformation efficiency data. (A-D) ComEC (EC) is labeled in gray. ComFA (FA) is labeled in blue. Arrows indicate relative strength of import, displayed as the DNA moving from left to right. (A) Wildtype condition. (B) $comFA^{\Delta}S1$ and combination mutants. It is unclear what the combination defect is, but the helicase and zinc finger functions have been disrupted. (C) DEAD-box helicase motif mutants. ComFA cannot move along the incoming DNA strand, but can bind to the naked ssDNA as it enters the cell. The muted ComFA surrounded by the dashed line represents a secondary binding event. The binding events would be expected to continue as the DNA moves into the cytoplasm. (D) $comFA^{4CS}$ mutants. The zinc finger mutants should maintain helicase activity, but the rate of DNA import would be reduced to a loss of localization, or interaction with other proteins required for transport.

4.5 Future directions and next steps

4.5.1 Further analysis of ComFA

The majority of the work I have presented here involves an *in vivo* analysis of mutations in ComFA. This leaves a lot that we still need to understand about how ComFA functions mechanistically in order to properly illustrate its role during DNA uptake. I will address a few of these experiments.

4.5.1.1 *In vivo* experiments

While conducting my analysis of ComFA I encountered a number of issues examining DNA uptake using methods developed for analysis in *B. subtilis* (48, 58). However, the Dubnau lab has managed to develop a protoplasting method which demonstrated that DNase-resistance is not a direct indicator of cytoplasmic transport (48). This technique may be adapted to measure the second step in internalization. Monitoring the second step will allow a more direct examination of the effects of ComFA mutation on DNA uptake during transformation.

Tracking internalization may also be possible using fluorescent microscopy, or flow cytometry, in which the substrate, or a secondary reporter is labeled such that transport can be measured rapidly and quantitatively. An important consideration for these types of techniques will be controlling any potential interference with the transport process caused by the chosen probe or label.

There also remains a great deal of work to be done in examining the localization of ComFA, and other components of the competence machinery. Much of the work extant in the literature has not commanded broad confidence due to the chosen method of examination, or issues with reproducibility, or demonstration of function (45, 50, 57, 59, 130, 131).

4.5.1.2 *In vitro* experiments

The proteins which constitute the competence machinery are very early in their biochemical study. None of them have been rigorously purified, and most of the structural or other biochemical information we have is based upon inferences based on homology or *in vivo* phenotype analysis.

In the case of ComFA, the work I performed here lays the foundation of intimate analysis of how the protein functions. The MBP-ComFA construct I have developed complements the null phenotype *in vivo* (See Figure 3.9). Furthermore, I have developed a method to remove contaminating DNA present during purification (See Figures 3.8, 3.12, & 3.13), and it is possible to remove the MBP tag using proteolytic cleavage (See Figure 3.10B), but more work is needed to improve the solubility of the cleaved protein.

The preliminary electrophoretic mobility shift assays (EMSAs) I performed suggest that ComFA binds single-stranded DNA (ssDNA) (See Figure 3.8B), however, a thorough analysis of substrate preference and analysis of helicase activity *in vitro* is needed to really determine if ComFA participates in melting the double helix of the incoming DNA, separating the

transforming and non-transforming strands.

4.6 Other members of the *comF* operon

ComFA is part of a tricistronic operon. Of the proteins in the operon, ComFA has the largest contribution to efficient transformation; however ComFB and ComFC do show minor contributions (58). ComFC is predicted to have a phosphoribosyl transferase function based on MotifScan analysis (113, 115), and is conserved in Gm^+ and Gm^- bacteria (58). ComFB and ComFC provide an opportunity to further study the requirement for metal co-factors during transformation and in the case of ComFC, possibly furthering our understanding of the fate of the DNA substrate during transformation.

4.6.0.3 Analysis of *comFC*

Of the other members of the *comF* operon ComFC would be the best protein to analyze to better understand transformation. Unlike ComFA, ComFC is conserved in both Gm^- and Gm^+ bacteria(58). It also contains a predicted four-cysteine metal binding motif, and a predicted phosphoribosyl transferase domain as well(58). Much like ComFA we understand that it has a marked influence on transformation efficiency, however we don't fully understand how it contributes to transformation. Though it is a much smaller protein, performing a biochemical analysis of ComFC would require solving a number of the obstacles I have encountered with ComFA. As such, it may now be easier to examine ComFC. Since ComFC

also contains a tetracysteine motif, some of the challenges to producing soluble protein, such as including zinc in the growth media during expression *in vitro* have been addressed.

4.7 Final Thoughts

4.7.1 Applications for natural transformation

Much of modern biological pursuits are dependent upon being able to perturb a system and isolating its components. For many organisms this is challenging or impossible due to barriers to HGT. By understanding which components are required for efficient transformation, it may be possible to mobilize genetic competence, enabling or improving the genetic tractability of previously impregnable biological systems. Similar approaches may be useful for developing methods for transfer of DNA to target specific cell-types in medical applications as well. ComFA is a very important piece of the machinery by increasing the rate of sequestration of transforming DNA from any degrading factors. Protecting your vector from degradation can be very important as maintaining DNA quality is critical when finding conditions for genetic transfer.

A

Appendix A: Analysis of expression of *comFA* and mutants at ectopic locus *yvbJ*

As an initial pass to determine the requirement for the DEAD-box helicase motifs in ComFA I made a series of constructs which expressed the ComFA and the mutant constructs under a P_{comF} promoter from the *yvbJ* locus, marked with a erythromycin resistance cassette. In this appendix I will describe these initial experiments and the findings from the ectopic experiment constructs.

A.1 Canonical DEAD-box mutations are not dominant

Analyzing the dominance of the DEAD-box mutation was performed while conducting preliminary analysis of the mutations. Mutant variants of ComFA were expressed from the *yvbJ* locus in *B. subtilis* from the endogenous *comF* promoter. The *comF* locus was left

intact. The meridiploids had an inherent 2-fold defect in transformation efficiency relative to the wild type. None of the strains showed a strong dominant-negative phenotype (Figure A.1). The system has been previously reported to be sensitive to protein levels (59), and while we do not observe a dominant-negative defect in any of the mutants in a meridiploid background, we do observe a defect in transformation efficiency in the meridiploid strains (Figure A.1).

Ectopic expression of *comFA* under a native P_{comF} promoter is unable to complement at wildtype levels (Figure A.2A) . It is possible that the failure to complement is the result of a genomic locus effect, the G to A transition mutation created during cloning, or some secondary effect from the antibiotic resistance cassette. This was not further investigated, as the endogenous system provided the most control in this experimental setup.

The meridiploid analysis findings do not agree with those previously reported by Londoño-Vallejo and Dubnau (49). There may be a number of reasons for the differences. For example the differences could arise from differences in the experimental set ups including differences in the media used to grow the cells to competence, or the markers for transformation. Some of the difference could also be attributed to differences in the construction of the strains, such as choice of expression loci. It is also unclear from the previous work what the level of variation was in the relative efficiencies observed.

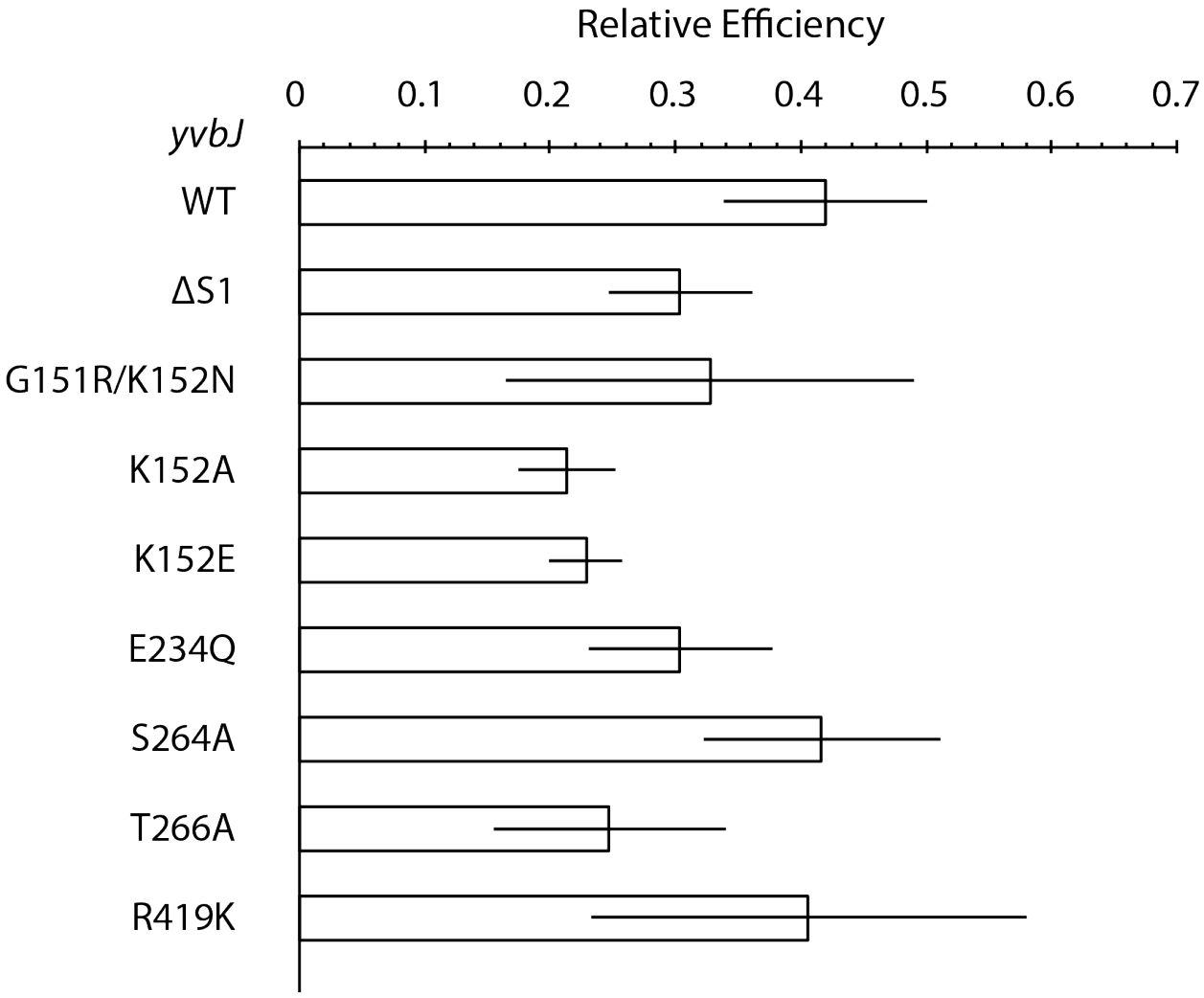

Figure A.1: DEAD-box mutations are not dominant - Transformation efficiency of meridiploid *B. subtilis* strains. Strains are wildtype at the *comF* locus. Mutation designations reflect mutations to ComFA integrated at *yvbJ*. Relative efficiencies are compared to wildtype *B. subtilis* PY79, with efficiency set to 1. Error bars are standard error. For all strains n=3.

A.2 Expression from *yvbJ* creates a competence defect

Typically, in *B. subtilis* genetic analyses are conducted by knocking out the gene of interest at the endogenous locus, and then examining complementation from an ectopic locus. There are a number of constructs which have been developed to express proteins from ectopic loci using inducible or endogenous promoters. The constructs are often marked with selectable genes which simplify making bacterial strains and verifying the maintenance of other insertions into the genome.

I began my analysis of ComFA in a similar manner. I made a *comFA* expression construct under the P_{comF} promoter that integrates into the genome at *yvbJ* to examine the contribution of ComFA to transformation efficiency. When ComFA was expressed from *yvbJ* as a meridiploid, I observed a mild transformation defect, of about 42 % of wildtype efficiency (Figure A.2A).

The mild defect was not entirely unexpected, as the transformation system had previously been reported to be sensitive to protein levels (58). When the endogenous *comF* locus is removed, I saw a much stronger defect in transformation efficiency, of approximately 4 % of wildtype. This additional defect may be the result of the loss of *comFB* and *comFC* which have both been shown to contribute to efficient transformation (58). The meridiploid and haploid mutational analysis experiments covered in this appendix are implemented with only ComFA being expressed from the *yvbJ* locus.

To determine if the defects observed were due to a protein level imbalance in the meridiploids, and/or a loss of *comFB* and *comFC* in the complementation haploid, I made new constructs which integrated the entire *comFABC* operon at *yvbJ*. Integrating the entire operon recovered some of the efficiency lost in the *comFA* only constructs, however, it was not sufficient to recapitulate the efficiency observed in the wildtype (Figure A.2B). The remaining defect may result from locus-specific effects, or the result of the G to A transition introduced in the start codon of the *comFA* coding sequence. The ATG start codon is generally assumed to be stroger, or better recognized by the ribosome than other variants such as the GTG we find in *comFA* in *B. subtilis*. This could increase relative levels of ComFA translation in the ectopic constructs relative to wildtype *B. subtilis*.

Given the observation that the complementation from the *yvbJ* locus was incomplete I decided to examine the effects of mutations made at the endogenous locus. The alternative system would alleviate any possible effects resulting from ectopic expression, or the inclusion of an antibiotic cassette. The results of the endogenous site analysis is covered in Chapter 2.

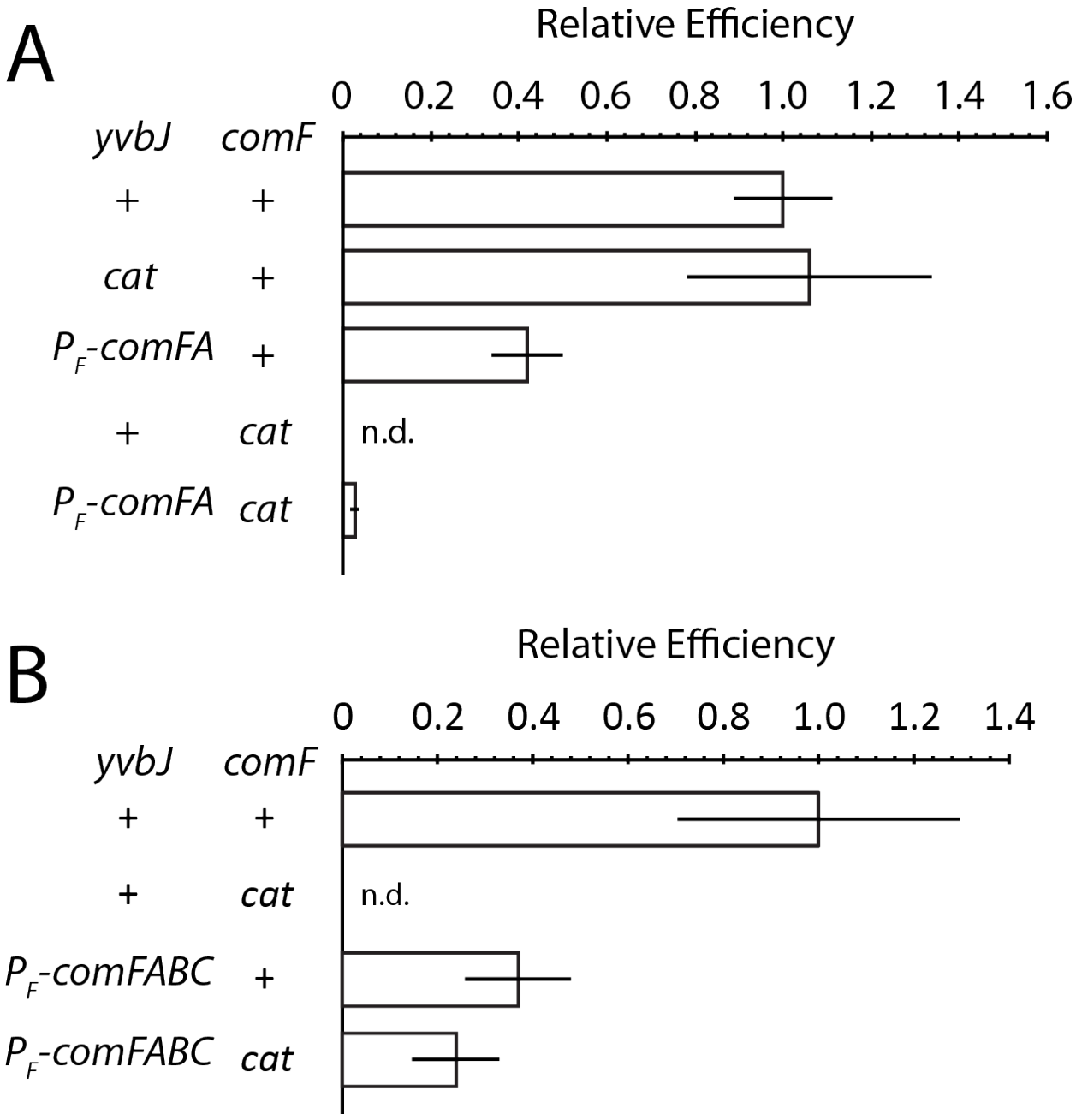

Figure A.2: *yvbJ::comFA* and *yvbJ::comFABC* expression do not achieve wild type transformation efficiency - (A and B) Relative efficiencies from complementation constructs with ComFA expressed from the *yvbJ* locus. P_F indicates P_{comF} promoter. n.d. = not determined. Error bars are standard error. (A) *yvbJ::comFA* constructs. (B) *yvbJ::comFABC* constructs. WT and *comF* meridiploid n = 2, *yvbJ::comFABC* n = 4.

A.3 Canonical DEAD-box motifs are required for ComFA function

Despite the above issues regarding complementation of the *yvbJ* ectopic expression constructs I performed a preliminary mutational analysis using mutations integrated at the ectopic location. To perform this analysis I replaced the *comF* locus in each of the meridiploid strains tested with an antibiotic resistance cassette. The resulting *B. subtilis* strains were haploid for *comFA*. The strains ectopically expressing *comFA* from the *yvbJ* locus showed a 10–fold defect from teh wildtype *B. subtilis* PY79 strain. All of the DEAD-box motif mutants (except $comFA^{S264A}$) showed at least an additional 10–fold defect in transformation efficiency from the *yvbJ::comFA* ectopic expression strain. Interestingly, the $comFA^{G151R/K152N}$ and $comFA^{E234Q}$ strains showed transformation efficiency rates comparable to the $comFA^{\Delta S1}$ in-frame deletion strain (Figure A.3)

A.4 Analysis of the ABC signature-like motif

Bioinformatic analysis of ComFA using MotifScan (113, 115) showed the presence of an accessory motif in a addition to the DEAD-box helicase motifs. An ABC signature-like motif was identified amino proximal to motif I (Figure A.4B). In ABC proteins the ABC Signature motif is involved in ATP-binding, and assists in dimerization. The ABC signature-like motif

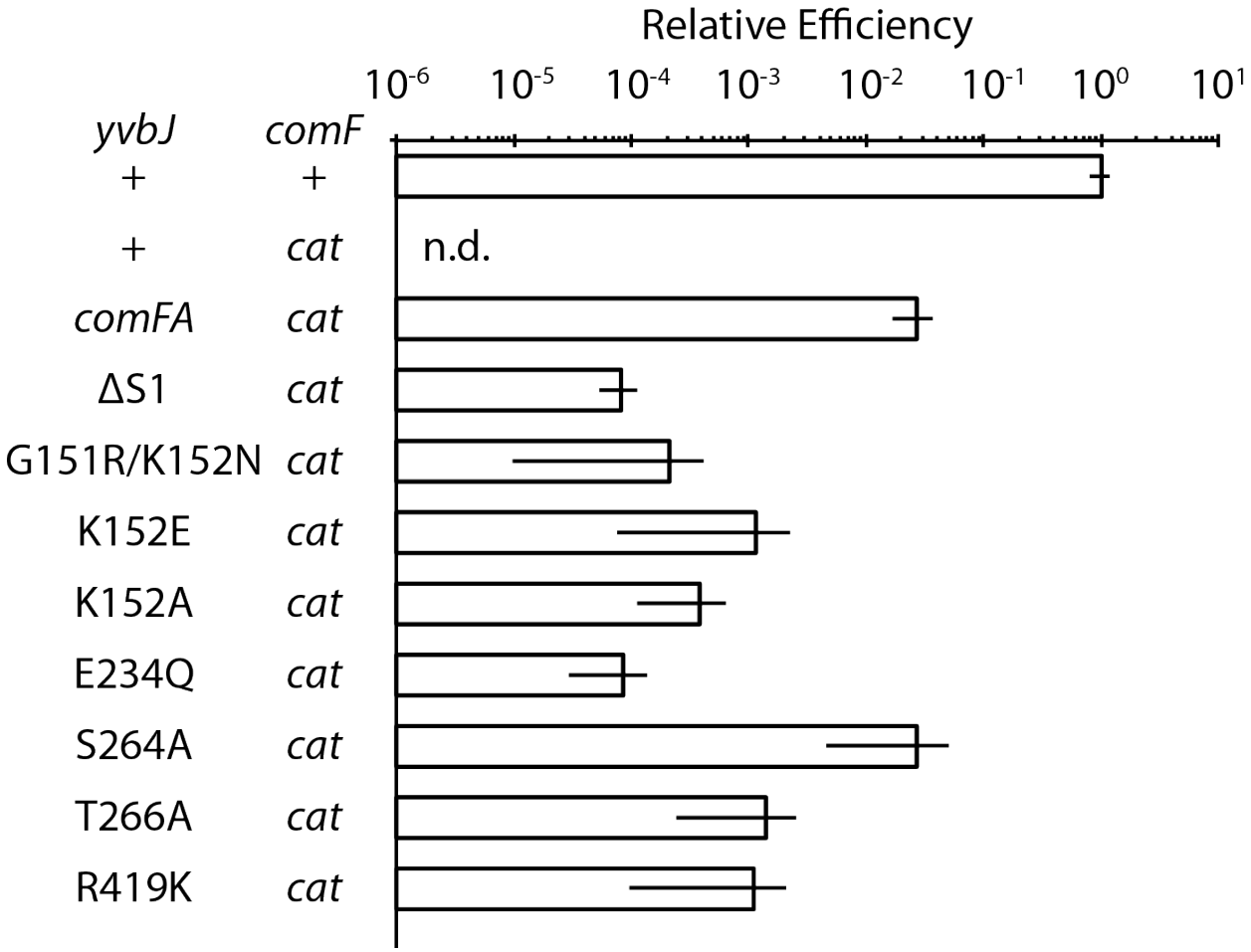

Figure A.3: Canonical DEAD-box motifs are required for ComFA function when expressed from ectopic locus - Relative transformation efficiencies for mutations made to *comFA* expressed from the *yvbJ* locus. (+) indicates that the locus is wild type. Mutations listed are made in the ComFA primary sequence. Error bars are standard error.

in ComFA is out of sequence with respect to the Walker A and Walker B ATP-binding and hydrolysis sites, but it is still possible that it participates in nucleotide binding. I also found that there is a conserved glutamine residue upstream of motif I in ComFA, which could indicate a DEAD-box Q motif (Figure A.4B and see Figure 2.2).

Figure A.4: Schematic of ComFA motif compared with DEAD Family and ABC Family proteins - (A and B) Zinc finger motif is indicated in silver, possible ABC signature motif is indicated in red, and DEAD family motifs are indicated in green. *Indicates residues mutated in transformation efficiency analysis(A) Enlarged region is removed in ΔS1 in-frame deletion. (B)Alignment of ComFA with DEAD family and ABC family conserved motifs.

To test these possible motifs I made three ComFA mutant constructs: *comFA*^{S122N}, *comFA*^{G124D}, and *comFA*^{Q125A}. The first two mutants would test the contribution of the ABC signature-like motif, and were based upon mutation that disrupt function in ABC transporters (150, 151). The third mutation was made to test the Q-motif, and is based upon the conservation observed in ComFA homologs, and DEAD-box helicases (see Figure 2.2).

The mutations in the ABC signature-like motif had no effect on transformation efficiency when introduced at the *comF* locus. Interestingly, when expressed as part of mutant *comFA* alleles from *yvbJ* in the absence of *comFB* and *comFC*, they both showed mild transformation efficiency defects (Figure A.5). The possible Q-motif conserved glutamine mutant was only tested as a mutation at the *comF* locus. The mutation caused a mild transformation efficiency defect on the order of about 2-fold (Figure A.5).

The results suggest that the ABC signature-like motif is more likely to be a variant of a Q-motif-like sequence (see Figures 2.2 & 3.1). The S122 and G124 appear to be dispensable, at least in the presence of ComFB and ComFC (Figure A.5). The requirement for Q125, but not for S122 or G124 is consistent with the ABC signature-like motif being a Q-motif variant. It is unclear what the cause of the defect observed in the *yvbJ::comFA* expression strains. The possible Q-motif maybe be functional, however, additional data is needed, as there is a nearby glutamine that could be substituting to maintain some activity. Also, the DEAD/H-box helicases that have been examined by mutational analysis do not all share the

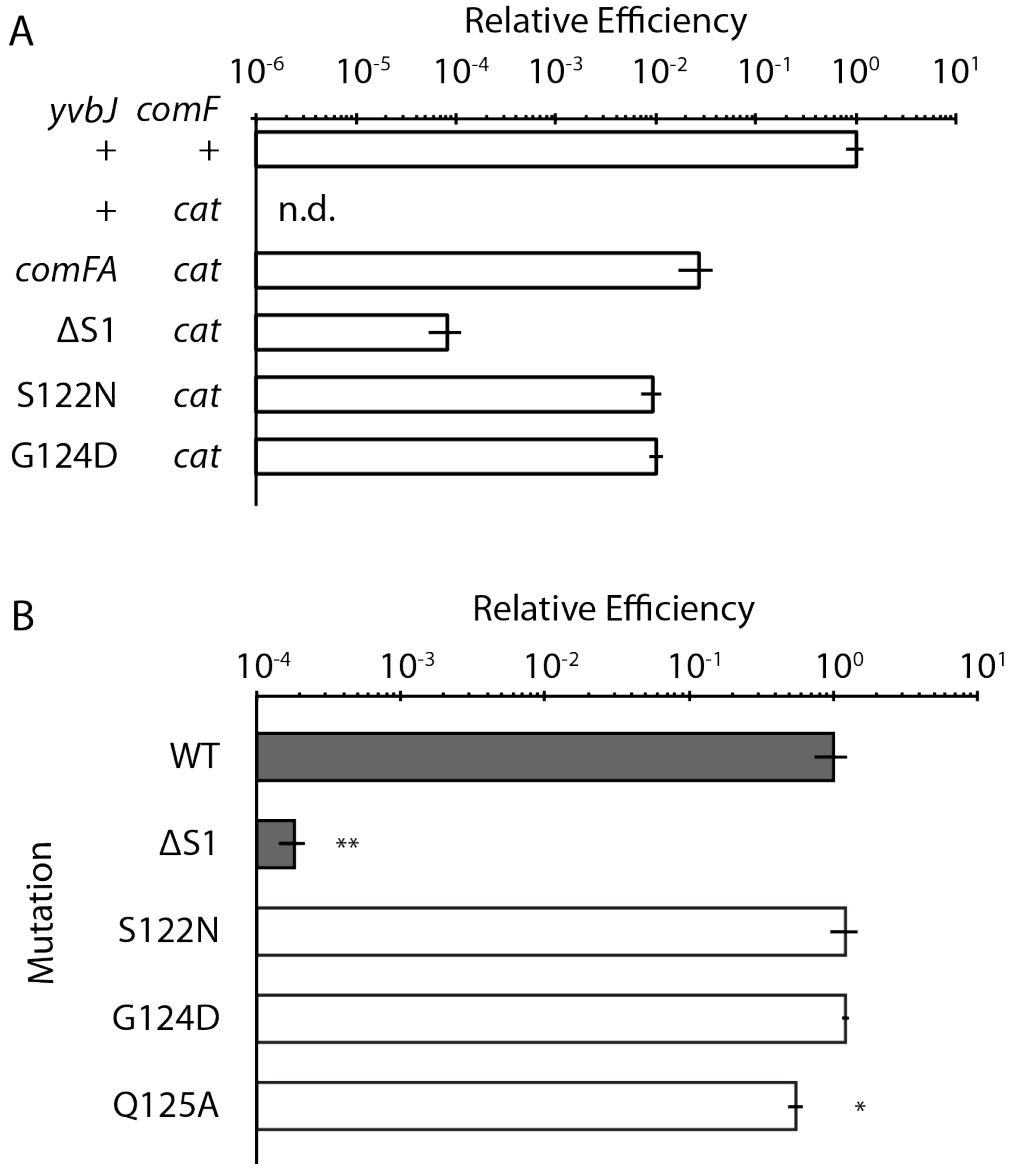

Figure A.5: The ABC signature-like motif is not required for efficient transformation - Transformation efficiency for mutations of ComFA ABC signature-like and DEAD helicase Q-like motifs. (A and B) All efficiency rates are normalized to wild type. Error bars are standard error. (A) (+) indicates that the locus is wild type. Mutations listed are made in the ComFA primary sequence. n.d. = not determined. Wildtype n=19, *yvbJ::comFA* n=16, ΔS1 n=4 S122N and G124D mutants n = 2 (B) Limit of detection for the assay is 0.5 transformants per CFU per μg of genomic DNA. The relative efficiency axis is a $\log_{10}$ scale. WT n=19, ΔS1, S122N, and G124D mutants n=5, Q125A mutant n=4. *p=0.005, **p<0.0001.

same phenotype when the Q-motif is mutated (152). The mild phenotype may indicate that the adenine recognition is not sufficiently perturbed, as the neighboring glutamine, Q126 could compensate for the loss of Q125. The mild phenotype may also suggest that the motif may serve a different function in ComFA, or ComFA is not strictly an ATPase, as helicases lacking the Q-motif sometimes utilize other nucleotides for energy (89, 153). Structural data with nucleotide bound would clarify this point. Determining which nucleotides can serve as energy sources for ComFA would also help to determine if the Q-motif-like sequence is a Q-motif variant or not.

A.5 Conclusion

The mutational analysis of ComFA from the *yvbJ* ectopic locus did provide some useful preliminary data, however it did create a number of issues as well. Expression of neither *comFA* alone, nor the *comF* operon from the *yvbJ* locus was sufficient to complement wild type transformation efficiency. There are a number of potential causes for the inherent defect including so called "site-specific" effects which could be alleviated by expressing the proteins from a different location, such as the *amyE* locus used by Londoño-Vallejo and Dubnau previously (49). The defect may also be caused by the use of the antibiotic cassette which marks the insertion in *yvbJ*. There may be some interference from the constitutively expressed resistance gene, and the induction of the *comF* protein products. Each of these potential scenarios are pretty straightforward to test and examine. Determining the cause of

the differences may be important to address when considering the minimal machinery, and possibly mobilizing the transformation machinery to other organisms.

One result that was particularly interesting however, was the change in the defect observed for the ABC signature-like motif mutants. The $comFA^{S122N}$, and $comFA^{G124D}$ mutants showed transformation efficiency defects when expressed from the *yvbJ* locus, relative to the wild type *comFA* gene expressed in the same manner (Figure A.5A), yet these defects were alleviated when the mutants were expressed from the *comF* locus (Figure A.5B). It is not clear whether that is the result of the absence of *comFB* and *comFC*, which would suggest an interaction between the proteins, missing from our models of the machinery, or whether there is something more mysterious about the mutations in the context of the *yvbJ* locus.

Resolving the ectopic expression defects will be important to allow further meridiploid analyses of members of the *comF* operon. The mutations I tested did not show a dominant negative phenotype, which conflicts with previous reports for the effects of mutations in the DEAD-box protein motif I present in ComFA (49). Reconciling the differences there may likely require a different approach, or moving the construct to a new locus. There also may be mutations in *comFA* which behave in a dominant-negative manner, and having a reliable method of expressing ComFA alleles from a second locus in the *B. subtilis* genome will be valuable in examining those mutant alleles.

The differences in the results I found in Chapter 2 and this appendix suggest that there is a great deal of value to expressing mutant alleles from their endogenous loci. The use

of ectopic loci arises from a number of historical challenges, which are less profound with advancements in biotechnology. Overtime the difficulty and expense of procedures such as PCR and sequencing have decreased, making markerless constructs more feasible. Also, prior to the development of isothermal assembly (94) using ectopic loci simplified strain construction as modular plasmids could be used to quickly make changes.

The double-crossover marker replacement integration is often used to increase strain generation rates, as of the target ectopic loci is marked with a selective marker, and selective-marker replacement can be used as a proxy for integration. The selective markers also allow the rapid transfer of alleles between strains as transfer of the relevant region of genomic DNA can be tracked by the transfer of a selective marker linked to your gene of interest.

In the markerless system I used in Chapters 2 & 3 (97) each mutation must be inserted in series, unless they are sufficiently clustered in the DNA sequence. For example, all four of the cysteine to serine mutations made in the $comFA^{4CS}$ were transfered all at once since they were very close together, however, a separate transformation was required to make the $comFA^{4CS}$ combinations with the mutations in the canonical DEAD-box motifs. Each of those constructs must be screened by sequencing in the final step of constructions to determine which isolated contain the desired mutation. The genomic DNA from markerless strains also cannot be used to directly make new strains as there is no way to efficiently determine which isolates acquired the allele of interest following transformation.

The greatest advantage of the marked, ectopic expression system is the amount of time

required to construct each strain. In the markerless system, the time from preparing your recipient strain to the time of having sequence-verified mutants is about 10 days (the timeline can be shorter if making insertions or deletions). The marker-replacement method takes about half of that time. There are some improvements to the plasmid construct that could decrease the timeline for constructing markerless strain, such as using a bacteriocidal antibiotic resistance selective marker, rather than a bacteriostatic antibiotic resistance selective marker. The change would decrease the time by removing at least on step in which you need to isolate resistant cells from non-growing antibiotic sensitive cells. However, even with that improvement the marker-replacement method simplifies strain verification since you have the selective marker as a proxy.

I would argue though that the changes I found in transformation efficiency assays for the *comF* expressed mutants (Figure A.5B & see Figure 2.3) and the *yvbJ* expressed mutants (Figures A.3 & A.5A) should bring some pause to analyses performed using alleles expressed from ectopic loci. Even accounting for the defect inherent in the *yvbJ::comFA* construct, the behavior of the mutations relative to each other differs between those experiments. For example, the canonical DEAD-box mutations are equivalent to the $comFA^{\Delta S1}$ in-frame deletion in the *yvbJ* ectopic expression system (Figure A.3), but were 100–fold more efficient in the *comF* endogenous expression system (see Figure 2.3). Furthermore, the $comFA^{S122N}$ and $comFA^{G124D}$ only showed a defect relative to wildtype in the *yvbJ* ectopic expression system (Figure A.5). The interpretation of the results discussed in Chapters 2 & 3 would have

been very different if all of the mutants and combinations were equivalent to $comFA^{\Delta S1}$. The similarity between the mutants I observed in the ectopic expression system could possibly explain the observations made in (49).

A.6 Materials and methods

A.6.1 Strains and growth conditions

All *B. subtilis strains* were derived from the prototrophic strain PY79 (92). *B. subtilis* were grown in Luria-Bertani (LB) broth or on LB plates fortified with 1.5 % Bacto agar at 24 °C or 37 °C as appropriate. 10x modified competence (MC) medium was made as described in (93). Competent cells were grown in 1x MC supplemented with 0.3 % (v/v) 1 M $MgSO_4$. When appropriate, antibiotics were included at the following concentrations: 5 μg/ml chloramphenicol (Cm_5), 100 μg/ml spectinomycin ($Spec_{100}$), and 1 μg/ml erythromycin plus 25 μg/ml lincomycin (*mls*).

A.6.2 Plasmid construction

Plasmids used in this work are listed in Table A.2. Oligonucleotides used in this work are listed in Table A.3

pSC002[h_6-*comFA*^{K152A}] was generated by site-directed mutagenesis of pBB031 using oSC002.

pSC004[h_6-*comFA*^{S264A}] was generated by site-directed mutagenesis of pBB031 using oSC009.

pSC005[h_6-*comFA*^{T266A}] was generated by site-directed mutagenesis of pBB031 using oSC010.

pSC015[h_6-*comFA*^{R419K}] was generated by site-directed mutagenesis of pBB031 using oSC024.

pSC018[*malE-comFA*^{K152A}] was generated by a two-way ligation between an *Nde*I-*Bam*HI PCR product containing *comFA*^{K152A} amplified from pSC002 using oSC044 and oSC061into pMAL-c5E cut with *Nde*I and *Bam*HI.

pSC026[h_6-*comFA*$^{G151R/K152N}$] See Chapter 2 Plasmid construction.

pSC036[*yvbJ::*P_{comF}*-comFA (erm)*] See Chapter 2 Plasmid construction.

pSC048[P_{comF}*-comFA* in pBlueScript SK(+)] See Chapter 2 Plasmid construction.

pSC051[P_{comF}-*comFA*^{T266A}] See Chapter 2 Plasmid construction.

pSC052[P_{comF}-$comFA^{R419K}$] was generated by a two-way ligation between an *Nde*I-*Bam*HI PCR product containing $comFA^{R419K}$ amplified from pSC015 using oSC044 and oSC061 into pSC048 cut with *Nde*I and *Bam*HI.

pSC056[P_{comF}-$comFA^{K152E}$] See Chapter 2 Plasmid construction.

pSC058[P_{comF}-$comFA^{G151R/K152N}$] See Chapter 2 Plasmid construction.

pSC059[*yvbJ*::P_{comF}-$comFA^{T266A}$ *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{T266A}$ from pSC051 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC060[*yvbJ*::P_{comF}-$comFA^{R419K}$ *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{R419K}$ from pSC052 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC064[*yvbJ*::P_{comF}-$comFA^{K152E}$ *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{K152E}$ from pSC056 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC066[*yvbJ::*P_{comF}-*comFA*$^{G151R/K152N}$ *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-*comFA*$^{G151R/K152N}$ from pSC058 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC067[P_{comF}-*comFA*^{S264A}] See Chapter 2 Plasmid construction.

pSC068[P_{comF}-*comFA*^{K152A}] See Chapter 2 Plasmid construction.

pSC069[P_{comF}-*comFA*^{E234Q}] See Chapter 2 Plasmid construction.

pSC073[P_{comF}-*comFA*$^{\Delta S1}$] See Chapter 2 Plasmid Construction.

pSC075[*yvbJ::*P_{comF}-*comFA*^{S264A} *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-*comFA*^{S264A} from pSC067 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC076[*yvbJ::*P_{comF}-*comFA*^{K152A} *(erm)*] See Chapter 2 Plasmid construction.

pSC077[*yvbJ::*P_{comF}-*comFA*^{E234Q} *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-*comFA*^{K152A} from pSC069 into pBB268 cut with *Eco*RI

and *Bam*HI.

pSC081[*yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$ *(erm)*] See Chapter 2 Plasmid construction.

pSC104[*comF::cat*] See Chapter 2 Plasmid Construction.

pSC116[*yvbJ::comFBC (erm)*] was generated by a two-way ligation with an *Eco*RI-*Bam*HI PCR product containing *comFBC* amplified from *B. subtilis* genomic DNA using oSC191 and oSC192, into pBB268 cut with *Eco*RI and *Bam*HI.

pSC117[*yvbJ::comFBC (erm)*] was generated by site-directed mutagenesis of pSC116 with oSC216 to remove a *Bgl*II site.

pSC172[*yvbJ::*P_{comF}*-comFABC*] was generated by a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}*-comFA* into pSC117 cut with *Eco*RI and *Bgl*II .

pSC192[P_{comF}*-comFA*^{S122N}] was generated by site-directed mutagenesis of pSC048 using oSC006.

pSC193[P_{comF}*-comFA*^{G124D}] was generated by site-directed mutagenesis of pSC048 using

oSC007

pSC221[P_{comF}-$comFA^{S122N}$ in pMiniMAD2] was generated by two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{S122N}$ from pSC192 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC222[P_{comF}-$comFA^{G124D}$ in pMiniMAD2] was generated by two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}-$comFA^{G124D}$ from pSC193 into pMiniMAD2 cut with *Eco*RI and *Bam*HI.

pSC236[$comFA^{1\text{-}259}$ in pMiniMAD2] See Chapter 3 Plasmid construction.

pSC240[$comFA^{\Delta S1}$ in pMiniMAD2] See Chapter 2 Plasmid construction.

pSC299[$comFA^{1\text{-}259,Q125A}$ in pMiniMAD2] was generated by site directed mutagenesis of pSC236 using oSC433.

A.6.3 *B. subtilis* strain construction

A.6.3.1 Meridiploid strains

Mutant versions of *comFA* were inserted at the *yvbJ* locus, marked with an erythromycin resistance marker. The resultant *B. subtilis* strain (bSC042) was used a host for the transformations. The host strain was grown at 37 °C for 4 hours in 1x MC. One colony per 1 ml of 1x MC was used for initial inoculum. While the cells were grown up, the relevant plasmids to be used for integration into the *yvbJ* locus were linearized using ScaI-HF endonuclease (New England Biolabs). At 4 hours 2 μl and 18 μl of linearized vector were added to 200 μl of 1x MC culture in 13 mm borosilicate glass tubes and grown for an addition 2 hours at 37 °C. After the additional incubation the transformation cultures were plated on LB/*mls* agar plates and grown overnight at 37 °C. Following the overnight incubation 8 colonies were chosen and struck out for single colonies on LB/*mls* plates and incubated overnight at 37 °C. Proper integration was tested by marker replacement. One colony from each of the 8 isolates chosen was picked and patched on LB/*mls* and LB/Cm_5 agar plates. Patches that showed *mls* resistance and chloramphenicol sensitivity were chosen for use in experiments. Glycerol stocks were made from chosen strains and stored at –80 °C.

A.6.3.2 Haploid *yvbJ* expression strains

Strains were made *comF* by transformation with genomic DNA from a *comF::cat* strain (bSC049). The *comF::cat* strain was made by the same method as the production of meridiploid strains, except pSC104 was used for integration, *B. subtilis* PY79 was used as host strain, and strains selected for chloramphenicol resistance. Replacement verified by PCR amplification of locus using oSC085 and oSC086. Genomic DNA was extracted from the resultant strain (bSC049) and used for creation of haploid strains. Each host strain was grown at 37 °C for 4 hours in 1x MC. One colony per 1 ml of 1x MC was used for initial inoculum. Dilutions of the genomic DNA bSC049 made at 1:20 and 1:400 into MiliQ H_2O. At 4 hours 2 μl of each dilution was added to 200 μl of 1x MC culture and grown for an addition 2 hours at 37 °C. After the additional incubation the transformation cultures were plated on LB/Cm_5 agar plates and grown overnight at 37 °C. Following the overnight incubation 8 colonies were chosen and struck out for single colonies on LB/Cm_5 plates and incubated overnight at 37 °C. Retention of the original integration was verified by antibiotic selection.

A.6.4 Transformation efficiency

Transformation efficiency experiments were performed as in Chapter 2. The tested strains were streaked out on LB, LB/Cm_5, or LB/*mls* plates depending on the resistance carried

rather than LB alone, and were grown overnight in the presence of antibiotics when appropriate.

Table A.1: Strains used in Appendix A

Name	Features	Reference
B. subtilis PY79		(92)
bBB364	*yvbJ::cat*	B.M. Burton
bSC017	*yvbJ::*P_{comF}*-comFA*^{T266A} *(erm)*	This work
bSC018	*yvbJ::*P_{comF}*-comFA*^{R419K} *(erm)*	This work
bSC022	*yvbJ::*P_{comF}*-comFA*^{K152E} *(erm)*	This work
bSC023	*yvbJ::*P_{comF}*-comFA*$^{G151R/K152N}$ *(erm)*	This work
bSC025	*yvbJ::*P_{comF}*-comFA*^{S264A} *(erm)*	This work
bSC026	*yvbJ::*P_{comF}*-comFA*^{K152A} *(erm)*	This work
bSC027	*yvbJ::*P_{comF}*-comFA*^{E234Q} *(erm)*	This work
bSC031	*yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$ *(erm)*	This work
bSC032	*yvbJ::*P_{comF}*-comFA (erm)*	This work
bSC049	*comF::cat*	This work
bSC105	*yvbJ::*P_{comF}*-comFABC (erm)*	This work
bSC106	*yvbJ::*P_{comF}*-comFABC (erm)*	This work
bSC107	*comF::cat*; *yvbJ::*P_{comF}*-comFABC (erm)*, derived from bSC105	This work
bSC108	*comF::cat*; *yvbJ::*P_{comF}*-comFABC (erm)*, derived from bSC105	This work
bSC109	*comF::cat*; *yvbJ::*P_{comF}*-comFABC (erm)*, derived from bSC106	This work
bSC110	*comF::cat*; *yvbJ::*P_{comF}*-comFABC (erm)*, derived from bSC106	This work
bSC140	*comF::cat*; *yvbJ::*P_{comF}*-comFA (erm)*	This work
bSC148	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{T266A} *(erm)*	This work
bSC150	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{R419K} *(erm)*	This work
bSC152	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{K152E} *(erm)*	This work
bSC154	*comF::cat*; *yvbJ::*P_{comF}*-comFA*$^{G151R/K152N}$ *(erm)*	This work
bSC156	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{S264A} *(erm)*	This work
bSC158	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{K152A} *(erm)*	This work
bSC160	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{E234Q} *(erm)*	This work
bSC166	*comF::cat*; *yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$ *(erm)*	This work
bSC168	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C60S} *(erm)*	This work
bSC170	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C63S} *(erm)*	This work
bSC172	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C84S} *(erm)*	This work

Continued on next page

Table A.1 – *Continued from previous page*

Name	Features	Reference
bSC174	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{C87S} *(erm)*	This work
bSC176	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{S122N} *(erm)*	This work
bSC178	*comF::cat*; *yvbJ::*P_{comF}*-comFA*^{G124D} *(erm)*	This work
bSC194	*comFA*^{G124D}	This work
bSC203	*comFA*^{S122N}	This work
bSC299	*comFA*^{Q125A}	This work

Table A.2: DNA & Plasmids used in Appendix A

Name	Features	Reference
Genomic DNA		
gSC018	*yhdGH::spec*	
gSC019	*comF::cat*	This work
Plasmids		
pBlueScript SK (+)	*bla*	
pBB268	*yvbJ::erm bla*	B.M. Burton
pMAL-c5E	*malE bla*	New England Biolabs
pMiniMAD2	*ori*^{BsTs} *bla erm*	(97)
pBB031	*comFA kan*	B.M. Burton
pSC002	*comFA*^{K152A}, derived from pBB031	This work
pSC004	*comFA*^{S264A}, derived from pBB031	This work
pSC005	*comFA*^{T266A}, derived from pBB031	This work
pSC015	*comFA*^{R419K}, derived from pBB031	This work
pSC018	*malE-comFA*^{K152A}	This work
pSC026	*comFA*$^{G151R/K152N}$, derived from pBB031	This work
pSC036	*yvbJ::*P_{comF}*-comFA*	This work
pSC048	pBlueScript SK(+) with P_{comF}*-comFA* from pSC036	This work
pSC051	*comFA*^{T266A}, derived from pSC048	This work

Continued on next page

Table A.2 – *Continued from previous page*

Name	Features	Reference
pSC052	$comFA^{R419K}$, derived from pSC048 with insert from pSC015	This work
pSC056	$comFA^{K152E}$, derived from pSC048	This work
pSC058	$comFA^{G151R/K152N}$, derived from pSC048, with insert from pSC026	This work
pSC059	$yvbJ::P_{comF}$-$comFA^{T266A}$, derived from pBB268 with insert from pSC051	This work
pSC060	$yvbJ::P_{comF}$-$comFA^{R419K}$, derived from pBB268 with insert from pSC052	This work
pSC064	$yvbJ::P_{comF}$-$comFA^{K152E}$, derived from pBB268 with insert from pSC056	This work
pSC066	$yvbJ::P_{comF}$-$comFA^{G151R/K152N}$, derived from pBB268 with insert from pSC058	This work
pSC067	P_{comF}-$comFA^{S264A}$, derived from pSC048 with insert from pSC004	This work
pSC068	$comFA^{K152A}$, derived from pSC048 with insert from pSC018	This work
pSC069	$comFA^{E234Q}$, derived from pSC048	This work
pSC073	$comFA^{\Delta S1}$, derived from SacI digest of P_{comF}-$comFA$ inserted into pUC19	This work
pSC075	$yvbJ::P_{comF}$-$comFA^{S264A}$, derived from pBB268 with insert from pSC067	This work
pSC076	$yvbJ::P_{comF}$-$comFA^{K152A}$, derived from pBB268 with insert from pSC068	This work
pSC077	$yvbJ::P_{comF}$-$comFA^{E234Q}$, derived from pBB268 with insert from pSC069	This work

Continued on next page

Table A.2 – *Continued from previous page*

Name	Features	Reference
pSC081	*yvbJ::*P_{comF}*-comFA*$^{\Delta S1}$, derived from pBB268 with insert from pSC073	This work
pSC104	*comF::cat*, derived from pSC047	This work
pSC116	*yvbJ::comFBC*, derived from pSC268	This work
pSC117	*yvbJ::comFBC*, derived from pSC116	
pSC172	*yvbJ::*P_{comF}*-comFABC*, derived from pSC117	
pSC192	*comFA*^{S122N}, derived from pSC048	This work
pSC193	*comFA*^{G124D}, derived from pSC048	This work
pSC221	pMiniMAD2 with insert from pSC192	This work
pSC222	pMiniMAD2 with insert from pSC193	This work
pSC236	pMiniMAD2 with *comFA*$^{1\text{-}259}$ and 424 bases upstream	This work
pSC240	*comFA*$^{\Delta S1}$, pMiniMAD2 with insert derived from pSC081 to otherwise resemble pSC236	This work
pSC299	*comFA*^{Q125A}, derived from pSC236	This work
pUC19	*bla*	

Table A.3: Oligonucleotides used in Appendix A

Name	Sequence
oSC002	CTGGGCGGTTTGCGGCGCTGGC**GCT**ACA GAAATGCTGTTTCCTGGTATA
oSC008	GATGCAATCGATGTTATGATCATTGAT**C**A **G**GTTGACGCTTTTCCATATTCTGC
oSC009	CAGCACCCTCGTTTATTTA**GCG**GCAACAC CTCCTAAAGAATT

Continued on next page

Table A.3 – *Continued from previous page*

Name	Sequence
oSC010	CACCCTCGTTTATTTAAGTGCA**G**C**G**CCTC CTAAAGAATTAAAAAGAAAAGC
oSC024	CATATTCTTTATGCCGGCCGGTT**T**TTCCT GCAATTTGAACAAGTGCGCT
oSC044	GTTTACTTTAAGAAGGAGATATACCATGG GCAGCAGCC
oSC061	CGGGATCCTAGTCTGTACATTCAACTTTT GCTGCC
oSC069	CTGGGCGGTTTGCGGCGCT**C**GCAA**C**ACA GAAATGCTGTTTCCTGGTATAGAATC
oSC191	CCGGAATTCCGGAAGATCTATTTGAAAAC GTGGTACAGTATACTC
oSC192	CGGGATCCTTAGCTTCTGATCAAGGTAAA AGATG
oSC216	GCGTTGGCCGATTCATTAATGCA**A**ATCTC GATCCCGCGAAATTAATACG

[a]Bold-face indicates mutagenic residues.
[b]Underlines indicate restriction sites.

B

Appendix B: Discovery of zinc finger motif

B.1 Introduction

B.1.1 Localization of competence components

One of the on-going areas of debate for the competence field is the localization of components of the DNA uptake machinery. A number of studies have examined the localization of competence components (50, 130, 131), however, the resolution of these studies is insufficient to fully understand the dynamics of the system. One of the original goals of this work was to examine the dynamics of ComFA and the other DNA uptake components during transformation. To perform this analysis I constructed a ComFA-GFP translational fusion. The construct was integrated into the *B. subtilis* genome at the *yvbJ* locus, under the P_{comF} promoter.

B.1.2 ComFA fluorescent tagging

B.1.2.1 Carboxy–terminal GFP is cleaved from ComFA

When cells grown to competence were observed by fluorescence microscopy the expressed protein showed a mostly diffuse cytosolic localization, with a number of brighter foci that were also present (Figure B.1). Foci were not observed in all cells, as expected during competence. To verify that the localization observed was a reflection of ComFA localization, *B. subtilis* lysates were analyzed by Western blot, probing against GFP. In the analysis I found that the ComFA-GFP localization pattern observed in Figure B.1 was the result of two populations of proteins; the full-length translational fusion, and some free GFP (Figure B.2). There were no observed intermediate degradation products. Since there was a significant quantity of free GFP present in the cells examined by fluorescence microscopy, I decided to look into some alternatives to the ComFA–GFP construct. I constructed an N–terminal GFP-ComFA translations fusion, and a ComFA–tetracysteine (ComFA–TC) construct which could be paired with a biarsenal fluorescent dye. The GFP-ComFA fusion showed a similar localization pattern to the ComFA-GFP construct (Figure B.1). However, the construct did not complement a *comFA* null allele, and so was not studied further (Figure B.3).

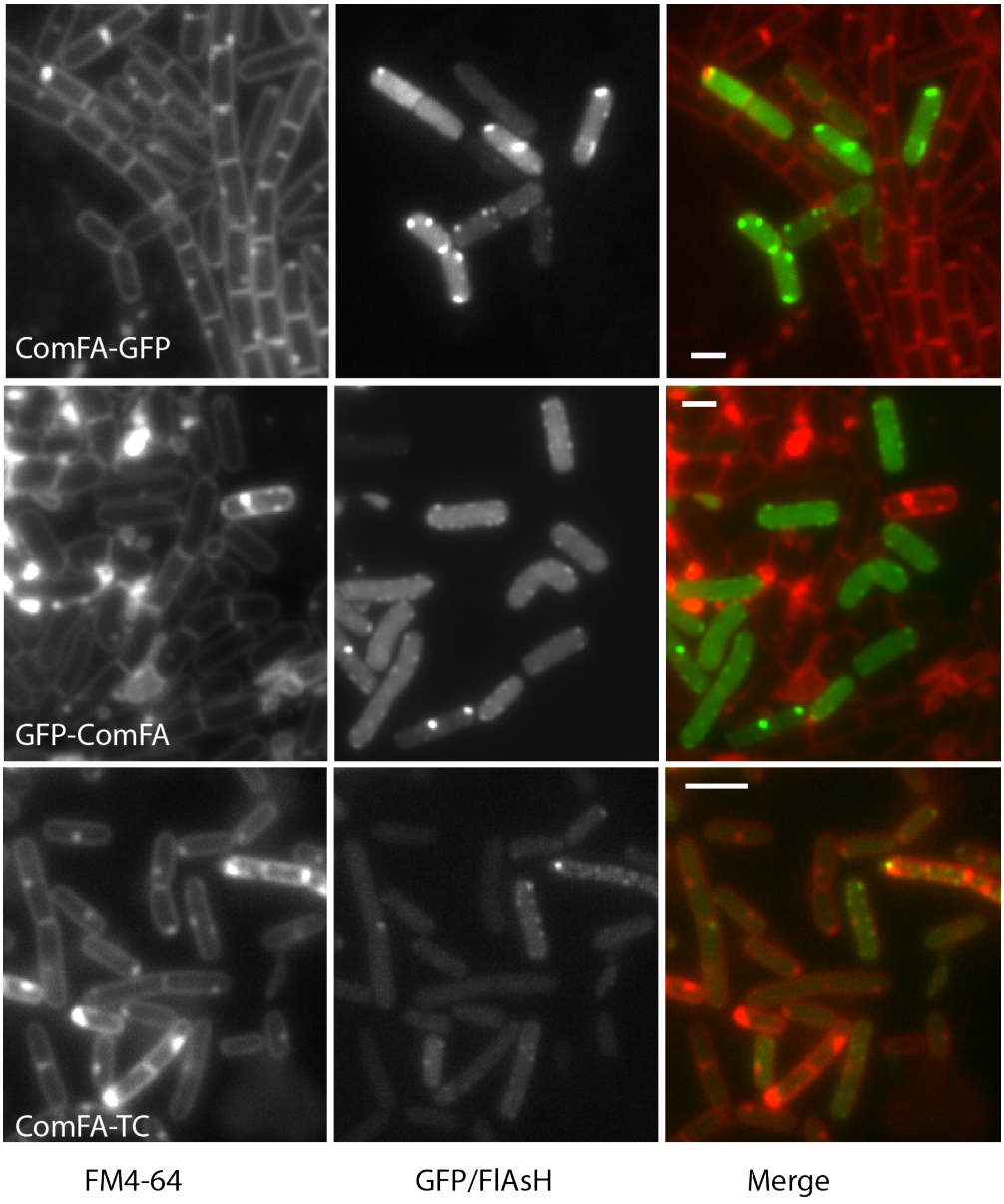

Figure B.1: Visualizing ComFA localization by fluorescence microscopy - Micrographs of *B. subtilis* cells expressing fluorescently-tagged ComFA. Each panel is labeled with the construct expressed. The first channel is the FM4-64 membrane stain. The second channel GFP fluorescence or FlAsH staining. The final panel is a merge of the previous two, with the FM4-64 false-colored red, and the FlAsH/GFP channel false-colored green.

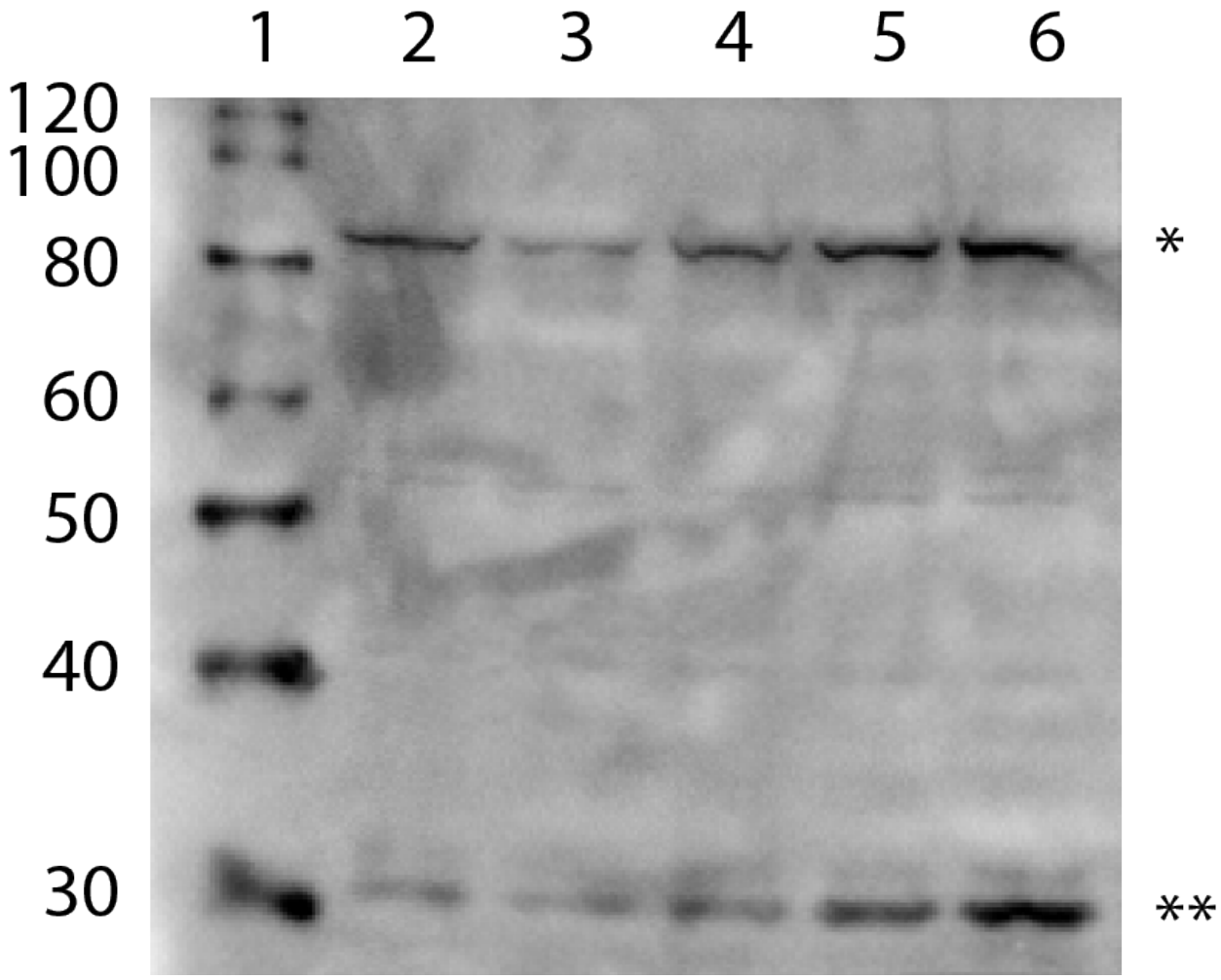

Figure B.2: Carboxy-terminal GFP is cleaved from ComFA - Western blot probing against GFP to determine if there are any degradation products in cells expressing ComFA-GFP. Samples are taken from a time course to examine ComFA-GFP expression. Samples were taken at 1–hour intervals beginning two hours post-dilution. Lane 1: Precision Plus Protein Standards, Lane 2: Pre-dilution, Lane 3: 2 hours post-dilution, Lane 4: 3 hours post-dilution, Lane 5: 4 hours post-dilution, Lane 6: 5 hours post-dilution. *ComFA–GFP, **GFP.

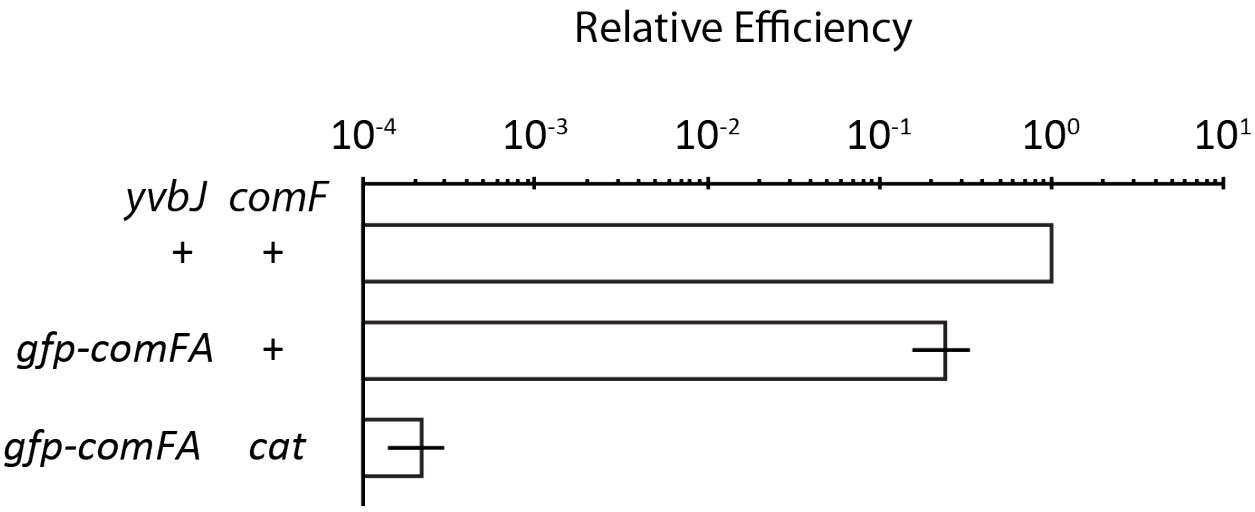

Figure B.3: GFP-ComFA does not complement - Relative transformation efficiency of *gfp–comFA* construct expressed from the *yvbJ* locus relative to wild type *B. subtilis* PY79. Wildtype n= 1, both *gfp–comFA* strains n=2. Error bars are standard error.

B.2 *comF* gene product binds FlAsH

While testing the biarsenal-tetracysteine FlAsH system I needed to verify that the signal observed was the result of the tagged proteins, and not some other gene product, or general background staining in *B. subtilis*. The system had not been tested in Gm^+ bacteria, even though it had been tested in a Gm^- bacterial system (154), so it was important to test for any intrinsic staining. I compared the *comFA-tc* strain to a strain lacking the tag, and a strain lacking the *comF* operon, when examined by fluorescence microscopy following FlAsH staining.

Initially, examining the micrographs suggested that there was some background in the untagged strain, however, there was some stronger signal associated with the tagged protein

(Figure B.4 inset). Histograms of pixel intensity showed something very interesting about the observed background staining. When comparing all three strains, I found that while there was a small change in the maximum intensity of pixels in the absence of the tag, a great deal of intensity was lost when the *comF* operon was ablated (Figure B.4).

The additional loss of staining signal suggested that the FlAsH cross-reactivity was the result of a *comF* gene product. I assumed at the time that ComFA was a good candidate, and one suggestion for pursing microscopy using the FlAsH system was to make a cysteine-less ComFA mutant which could then be tagged with the tetracysteine tag. An important component of the potential cysteine-less ComFA was that it needed to be functional. My approach to developing the cysteine-less allele was to systematically remove the cysteines, and test each intermediate to ensure that functionality was maintained. I began by examining the amino acid sequence of ComFA. I noticed that a tetracysteine tag-like sequence existed in the protein (Table B.1). Given the location in the sequence and the similarity to the TC tag I decided to begin mutagenizing the sequence at the two pairs of cysteines. Mutating the cysteines to serines created transformation defects, so the FlAsH tagging system was no investigated further as a method for examining the localization of ComFA. The results of my analysis of the TC tag-like region are covered in Chapter 3.

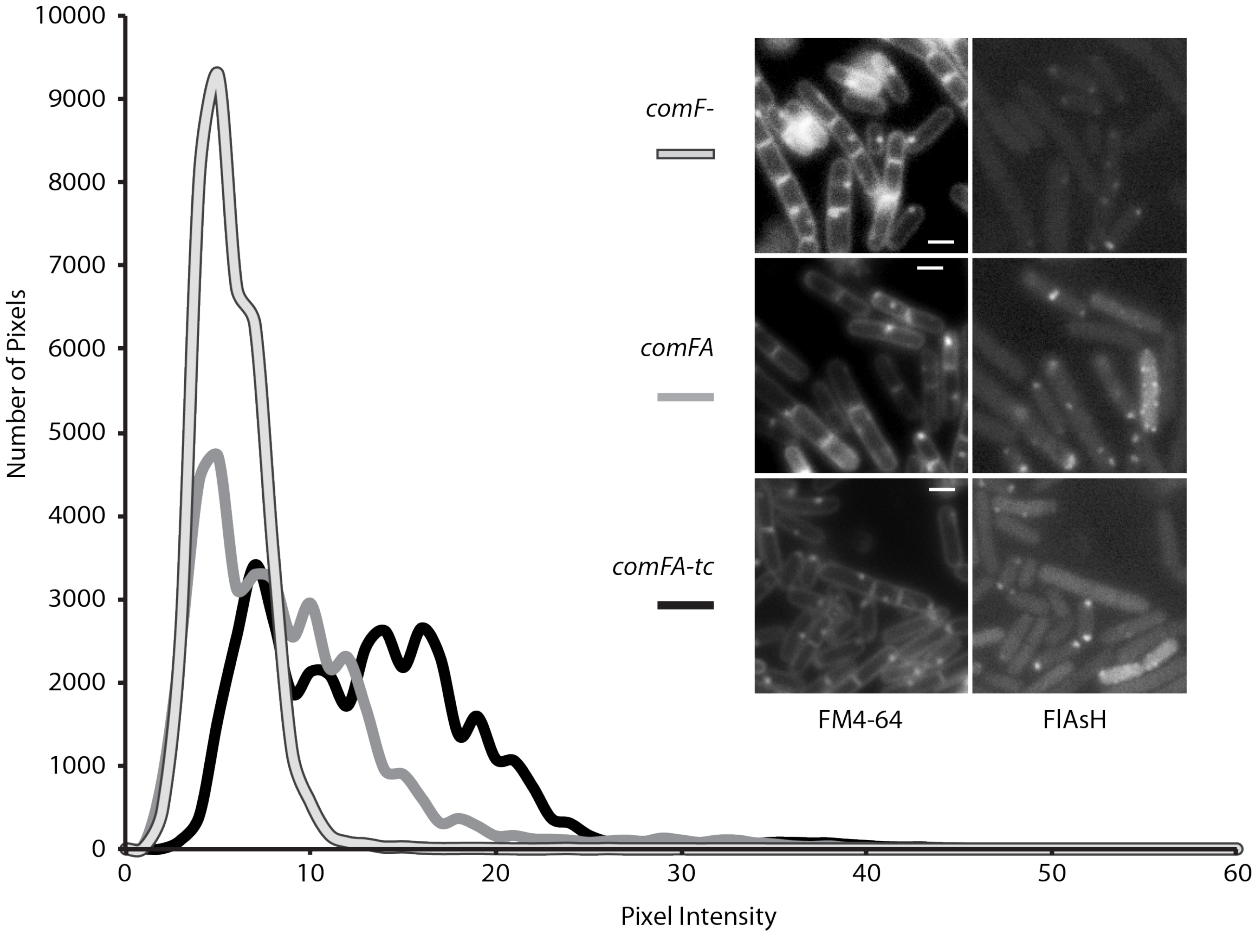

Figure B.4: FlAsH cross-reacts with *comF* gene product - Histogram of pixel intensities for the fields presented in the inset. The pixel intensity axis has been truncated to remove the higher values unoccupied by any pixels. Inset: Micrographs of FlAsH stained *B. subtilis* cells. The FM4-64 channel shows a membrane stain to indicate the cells. The FlAsH channel shows the FlAsH staining for each strain. The legend next to the inset corresponds with the classes compared in the histogram.

Table B.1: Comparison of ComFA N-terminal region and TC tag sequence

Name	Sequence
ComFA	50–SISINKRRYR**C**NR**C**GQTD QRYFSFYHSSGKKLY**C**RS**C**V MMGRVSEEVPLY–100
FlAsH	1–LEYKVVDAVNGKPIPNPLL GLDSTRTGAGG**CC**PG**CC**GG G–39

[a]Amino acid sequences. Numbers indicate first and last residue in sequence excerpt. Cysteines are bolded for emphasis.

B.3 Conclusion

Once it had been shown that the FlAsH biarsenal stain was able to cross react with what appeared to be ComFA the next step was to see if the cross-reactivity could be eliminated by possibly removing cysteines from ComFA. This led to the further examination of the ComFA amino acid sequence. When conducting this further examination I noticed there were two sets of paired cysteines early in the sequence, and decided to begin the process of making a cysteine-less ComFA by removing those cysteines. As discussed in Chapter 3, it turned out that those cysteines were required for proper function of the protein, and so no further work was conducted using the FlAsH system. Based on sequence analyses conducted in the literature (58) it is also possible that the FlAsH cross-reacted with ComFC, which as a tetracysteine motif as well.

B.4 Materials and methods

B.4.1 Strains and growth conditions

All *B. subtilis strains* were derived from the prototrophic strain PY79 (92). *B. subtilis* were grown in Luria-Bertani (LB) broth or on LB plates fortified with 1.5 % Bacto agar at 24 °C or 37 °C as appropriate. 10x modified competence (MC) medium was made as described in (93). Competent cells were grown in 1x MC supplemented with 0.3 % (v/v) 1 M $MgSO_4$. When appropriate, antibiotics were included at the following concentrations: 5 μg/ml chloramphenicol (Cm_5), and 1 μg/ml erythromycin plus 25 μg/ml lincomycin (*mls*).

B.4.2 Plasmid construction

pSC010 [gfp^{mut2b}] See Chapter 2 Plasmid construction.

pSC012 [*comFA*-gfp^{mut2b}] See Chapter 2 Plasmid construction.

pSC021 [pSC010 mutated to remove *Nde*I sites in gfp^{mut2b} and backbone] was generated by site-directed mutagenesis of pSC010 using oSC062 and oSC063 to remove *Nde*I sites from the gfp^{mut2b} coding sequence by silent mutation, and from the plasmid backbone.

pSC022 [*yvbJ::*P_{comF}*-comFA-*gfp^{mut2b} *(erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}*-comFA-*gfp^{mut2b} from pSC012 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC029 [P_{comF}*-*gfp^{mut2b}] was generated in a two-way ligation with an *Eco*RI-*Xho*I fragment containing P_{comF} from pSC010 into pSC021 cut with *Eco*RI and *Xho*I.

pSC031 [tetracysteine tag] was generated in a two-way ligation with an *Xho*I-*Bam*HI PCR product assembled using oSC077, oSC078, oSC079, and oSC080 into pSC021 cut with *Xho*I and *Bam*HI.

pSC033 [*comFA-tc*] was generated in a two-way ligation with an *Eco*RI-*Xho*I PCR product containing P_{comF}*-comFA* from pSC012 into pSC031 cut with *Eco*RI and *Xho*I.

pSC036 [*yvbJ::*P_{comF}*-comFA (erm)*] See Chapter 2 Plasmid construction.

pSC037 [*yvbJ::*P_{comF}*-comFA-tc (erm)*] was generated in a two-way ligation with an *Eco*RI-*Bam*HI fragment containing P_{comF}*-comFA-tc* from pSC012 into pBB268 cut with *Eco*RI and *Bam*HI.

pSC104 [*comF::cat*] See Chapter 2 Plasmid construction.

B.4.3 ComFA-GFP expression time course

bSC005 cells were streaked out from -80 °C storage onto LB/agar selective plates and incubated overnight at 37 °C. Overnight LB cultures were inoculated with a single colony and grown rolling at 24 °C. Cells grown in 1x MC for 4 hours at 37 °C. Following the 4-hour incubation the cells were diluted 1:1 000 in 1x MC media and grown for an additional 4 hours with samples taken at 1-hour intervals beginning at 2 hours of growth. Samples were taken for OD_{600}=1.5. Cells in samples were harvested by centrifugation, and cells suspended in 50 µl lysis buffer (20 mM Tris pH 7.5, 1 mM EDTA, 10 mM $MgCl_2$, 1 mg/ml lysozyme, 1 mM PMSF, 10 µg/ml DNAseI, 100 µg/ml RNAseA)*B. subtilis* samples were incubated at 37 °C for 10 minutes, and lysed by 1:1 dilution in 2x Sample buffer (250 mM Tris pH 6.8, 10 mM EDTA, 4 % (w/v) SDS, 20 % (v/v) glycerol, 0.3 % (w/v) bromophenol blue) and heated at 95 °C for 10 minutes. Sample run out on SDS-PAGE. ComFA-GFP cleavage analysed by western blot using anti-GFP polyclonal antibody.

B.4.4 Fluorescence microscopy

B.4.4.1 FlAsH

FlAsH obtained from Matthew Cabeen of the lab of Christine Jacobs-Wagner at Yale University. Cells grown in 1x MC for 4 hours at 37 °C. For microscopy 500 μl of each culture

transferred from growth tube to 1.5 ml microcentrifuge tubes. Cells pelleted by spinning down at 8 000 rpm (6 010 x g) for 1 minute. For use in staining 200 μl of spent media transferred to separate tube. Remaining spent media was aspirated from cells. Cell pellets were resuspended in 20 μl labeling reagent (5 μM FlAsH, 20.8 μM ethanedithiol (EDT), in spent media). Cells incubated at 37 °C for 30 minutes. Cells were washed 3x in 1 ml 1x PBS, 250 μM EDT to remove unbound FlAsH stain.Cells were then resuspended in 48 μl 1.04 mM EDT and spun again. Prior to imaging cells were resuspended in 1x PBS, 3 μg/ml FM4-64. Images captured for phase, green channel and red channel. FM 4-64 purchased as solid from Life Technologies, a 1 000x 1.5 mg/ml stock made, and used for dilution during staining.

B.4.4.2 Image analysis

Images acquired using Zeiss Imager M1 and Zeiss AxioVision software. Histograms taken from a representative field for each strain.

Table B.2: Strains used in Appendix B

Name	Features	Reference
bSC003	*amyE::*$P_{hyperspank}$*-comS (spec)*	
bSC004	*amyE::*$P_{hyperspank}$*-comS (spec); yvbJ::cat*, derived from bSC003	This work
bSC005	*amyE::*$P_{hyperspank}$*-comS (spec)*; *yvbJ::*P_{comF}*-comFA-gfp (erm)*	This work
bSC006	*amyE::*$P_{hyperspank}$*-comS (spec)*; *yvbJ::*P_{comF}*-comFA-gfp (erm)*	This work

Continued on next page

Table B.2 – *Continued from previous page*

Name	Features	Reference
bSC007	*comFA::Tn524 (erm)*	
bSC008	*amyE::*$P_{hyperspank}$*-comS (spec)*; *yvbJ::*P_{comF}*-comFA (erm)*	This work
bSC010	*amyE::*$P_{hyperspank}$*-comS (spec)*; *yvbJ::*P_{comF}*-comFA-tc (erm)*	This work
bSC050	*yvbJ::*P_{comF}*-gfp-comFA (erm)*	This work
bSC051	*yvbJ::*P_{comF}*-gfp-comFA (erm)*	This work
bSC055	*comF::cat*; *yvbJ::*P_{comF}*-gfp-comFA (erm)*	This work
bSC056	*comF::cat*; *yvbJ::*P_{comF}*-gfp-comFA (erm)*	This work

Table B.3: DNA & Plasmids used in Appendix B

Name	Features	Reference
pBB028	*bla cat*	B.M. Burton
pBB268	*bla yvbJ::erm*	B.M. Burton
pKL147	*bla spec gfp*^{mut2b}	(95)
pSC010	pKL147 with BamHI added	This work
pSC012	pSC010 with comFA inserted	This work
pSC014	intermediate in construction of pSC028, contains region downstream of *comFA*, derived from pSC009	This work
pSC021	pSC010 mutated to remove *Nde*I sites in *gfp* and backbone	This work
pSC022	*yvbJ::*P_{comF}*-comFA-gfp (erm)*, derived from pBB268 and pSC012	This work
pSC029	P_{comF} inserted into pSC021	This work
pSC031	tetracysteine tagging vector, derived from pSC021	This work
pSC033	*comFA-tc*, derived from pSC031	This work
pSC036	*yvbJ::*P_{comF}*-comFA (erm)*, derived from pBB268	
pSC037	*yvbJ::*P_{comF}*-comFA-tc (erm)*, derived from pBB268 and pSC033	This work
pSC104	*comF::cat*, derived from pSC047	This work

Table B.4: Oligonucleotides used in Appendix B

Name	Sequence
oSC012	GAACTATACAAATAAATGTCCAGAC GGATCCCTGCAGGCATGCAAGCTT
oSC062	GCTTTGCGAGATACCCAGATCA**C**ATGAAA CAGCATGACTTTTTCAAGAGTG
oSC063	CAGATTGTACTGAGAGTGCACCA**C**ATGCA AGGGTTTATTGTTTTCTAAAATCTG
oSC075	CGACCGATATATGGACCAGCTTCA**C**ATGG CCTGTACTTGCCAAGTATGC
oSC077	CCGCTCGAGTACAAAGTGGTTGATGCTGT TAACGGGAAGCCTATCCCTAACCC
oSC078	CCGGTACGCGTAGAATCGAGACCGAGGA GAGGGTTAGGGATAGGCTTCCCG
oSC079	GGTCTCGATTCTACGCGTACCGGTGCTGG TGGCTGTTGTCCTGGCTGTTGC
oSC080	CGGGATCCCTAGCCGCCACCGCAACAGCC AGGACAACAGCC
oSC193	ACATGCATGCATGATTCTGTTTTTATGCC GATATAATC
oSC194	GCTCTAGAGTTGCAGTCTTTAAACAATCT TAACCC

[a]Bold-face indicates mutagenic residues.
[b]Underlines indicate restriction sites.

References

[1] P. Schu. Vesicular protein transport. *Pharmacogenomics J*, 1:262–271, 2001. 2

[2] Susan R. Wente and Michael P. Rout. The nuclear pore complex and nuclear transport. *Cold Spring Harbor Perspectives in Biology*, 2(10), 2010. 2

[3] Tom A. Rapoport, Berit Jungnickel, and Ulrike Kutay. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annual Review of Biochemistry*, 65(1):271–303, 1996. PMID: 8811181. 2

[4] Tsai-Tien Tseng, Brett Tyler, and João Setubal. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, 9(Suppl 1):S2–S2, February 2009. 4

[5] S E Stachel and P C Zambryski. Agrobacterium tumefaciens and the susceptible plant cell: a novel adaptation of extracellular recognition and DNA conjugation. *Cell*, 1986. 4

[6] M G Lorenz and W Wackernagel. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological reviews*, 58(3):563–602, September 1994. 4, 17

[7] Hilde Steinmoen, Eivind Knutsen, and Leiv Håvarstein. Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. *Proceedings of the National Academy of Sciences*, 99(11):7681–7686, 2002. 4

[8] Hilde Steinmoen, Aina Teigen, and Leiv Sigve Håvarstein. Competence-induced cells of Streptococcus pneumoniae lyse competence-deficient cells of the same strain during cocultivation. *Journal of Bacteriology*, 185(24):7176–7183, December 2003. 4

[9] D Dubnau. DNA uptake in bacteria. *Annual review of microbiology*, 53:217–244, January 1999. 4

[10] Inês Chen and David Dubnau. DNA uptake during bacterial transformation. *Nature Reviews Microbiology*, 2(3):241–249, March 2004. 4, 9, 15

[11] D Moradigaravand and J Engelstädter. The evolution of natural competence: disentangling costs and benefits of sex in bacteria. *The American Naturalist*, January 2013. 4, 97

[12] Gregor Mendel. Versuche über Plflanzenhybriden . *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865*, Abhand-lungen:3–47, April 1866. 5

[13] Douglas Hanahan. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology*, 166(4):557–580, 1983. 5

[14] A Puyet, H Sandoval, P Lopez, A Aguilar, J F Martin, and M Espinosa. A simple medium for rapid regeneration of Bacillus subtilis protoplasts transformed with plasmid DNA. *FEMS microbiology letters*, 40(1):1–5, 1987. 5

[15] F Meinhardt, U Stahl, and W Ebeling. Highly efficient expression of homologous and heterologous genes in Bacillus megaterium - Springer. *Applied microbiology and biotechnology*, 1989. 5

[16] B Wang, X Yang, and R Wu. High-level production of the mouse epidermal growth factor in a Bacillus brevis expression system. *Protein expression and purification*, 4(3):223–231, June 1993. 5

[17] Shuang-En Chuang, Ann-Lii Chen, and Chih-Chiang Chao. Growth of E. coli at low temperature dramatically increases the transformation frequency by electroporation. *Nucleic acids research*, 23(9):1641, 1995. 5

[18] Roy Curtiss, III. Bacterial conjugation. *Annual Reviews in Microbiology*, 23(1):69–136, 1969. 5, 6

[19] Hideo Ikeda and Jun-ichi Tomizawa. Transducing fragments in generalized transduction by phage P1. *Journal of molecular biology*, 14(1):85–109, November 1965. 5

[20] D Dubnau, R Davidoff-Abelson, and I Smith. Transformation and transduction in Bacillus subtilis: evidence for separate modes of recombinant formation. *Journal of molecular biology*, 45(2):155–179, October 1969. 5

[21] Inês Chen, Peter J Christie, and David Dubnau. The ins and outs of DNA transfer in bacteria. *Science*, 310(5753):1456–1460, December 2005. 6, 8, 9, 15, 17

[22] Lars Andrup, Ole JØrgensen, Andrea Wilcks, Lasse Smidt, and Gert B Jensen. Mobilization of "Nonmobilizable" Plasmids by the Aggregation-Mediated Conjugation System of Bacillus thuringiensis. *Plasmid*, 36(2):75–85, 1996. 6

[23] Haruo Ozeki and Hideo Ikeda. Transduction mechanisms. *Annual Review of Genetics*, 2(1):245–278, 1968. 8

[24] Marion Dorer, Ilana Cohen, Tate Sessler, Jutta Fero, and Nina Salama. Natural competence promotes Helicobacter pylori chronic infection. *Infection and immunity*, 81(1):209–215, 2013. 8, 97

[25] Yao-Hui Sun, Rachel Exley, Yanwen Li, David Goulding, and Christoph Tang. Identification and characterization of genes required for competence in Neisseria meningitidis. *Journal of Bacteriology*, 187(9):3273–3276, 2005. 8, 93

[26] J Foldes and T A Trautner. Infectious DNA from a newly isolatedB. subtilis phage. *MGG Molecular & General Genetics*, 95(1):57–65, 1964. 8

[27] D M Green. Infectivity of DNA isolated from Bacillus subtilis bacteriophage, SP82. *Journal of molecular biology*, 10(3):438–451, 1964. 8

[28] O Ciferri, S Barlati, and J Lederberg. Uptake of synthetic polynucleotides by competent cells of Bacillus subtilis. *Journal of Bacteriology*, 104(2):684–688, November 1970. 8, 11, 94

[29] David B Danner, Robert A Deich, Kenneth L Sisco, and Hamilton O Smith. An eleven-base-pair sequence determines the specificity of DNA uptake in Haemophilus transformation. *Gene*, 11(3-4):311–318, November 1980. 8

[30] C Elkins, C E Thomas, H S Seifert, and P F Sparling. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. *Journal of Bacteriology*, 173(12):3911–3913, June 1991. 8

[31] Dawit Kidane, Begoña Carrasco, Candela Manfredi, Katharina Rothmaier, Silvia Ayora, Serkalem Tadesse, Juan C Alonso, and Peter L Graumann. Evidence for different pathways during horizontal gene transfer in competent Bacillus subtilis cells. *PLoS genetics*, 5(9):e1000630, September 2009. 8, 9, 13, 14, 70

[32] D Dubnau. The regulation of genetic competence in *Bacillus subtilis*. *Molecular Microbiology*, 5(1):11–18, January 1991. 9

[33] D Dubnau. Genetic competence in Bacillus subtilis. *Microbiological reviews*, 55(3):395–424, September 1991. 9, 10, 11, 12, 28

[34] D Dubnau. Binding and transport of transforming DNA by Bacillus subtilis: the role of type-IV pilin-like proteins–a review. *Gene*, 192(1):191–198, June 1997. 9

[35] Briana Burton and David Dubnau. Membrane-associated DNA transport machines. *Cold Spring Harbor Perspectives in Biology*, 2(7), 2010. 9, 18, 27

[36] D van Sinderen and G Venema. comK acts as an autoregulatory control switch in the signal transduction route to competence in Bacillus subtilis. *Journal of Bacteriology*, 176(18):5762–5770, September 1994. 9

[37] K Turgay, J Hahn, J Burghoorn, and D Dubnau. Competence in Bacillus subtilis is controlled by regulated proteolysis of a transcription factor. *The EMBO journal*, 17(22):6730–6738, November 1998. 9, 95

[38] Tolga Cağatay, Marc Turcotte, Michael B Elowitz, Jordi Garcia-Ojalvo, and Gürol M Süel. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139(3):512–522, September 2009. 9

[39] Florence Cahn and Maurice Fox. Fractionation of Transformable Bacteria from Competent Cultures of Bacillus subtilis on Renografin Gradients. *The Journal of Bacteriology*, 95(3):867, March 1968. 9

[40] C Hadden and E W Nester. Purification of competent cells in the Bacillus subtilis transformation system. *Journal of Bacteriology*, 95(3):876–885, March 1968. 9

[41] Leendert Hamoen, Gerard Venema, and Oscar Kuipers. Controlling competence in Bacillus subtilis: shared use of regulators. *Microbiology*, 149(1):9–17, 2003. 10

[42] David Dubnau and Carol Cirigliano. Fate of transforming DNA following uptake by competent Bacillus subtilis: IV. The endwise attachment and uptake of transforming DNA. *Journal of molecular biology*, 64(1):31–46, 1972. 11

[43] Berenike Maier, Inês Chen, David Dubnau, and Michael Sheetz. DNA transport into Bacillus subtilis requires proton motive force to generate large molecular forces. *Nature Structural & Molecular Biology*, 11(7):643–649, 2004. 11, 12, 16, 100

[44] R N Singh. Number of deoxyribonucleic acid uptake sites in competent cells of Bacillus subtilis. *Journal of Bacteriology*, 110(1):266–272, April 1972. 11

[45] R Provvedi, I Chen, and D Dubnau. NucA is required for DNA cleavage during transformation of Bacillus subtilis. *Molecular Microbiology*, 40(3):634–644, May 2001. 12, 15, 105

[46] James S Levine and Norman Strauss. Lag Period Characterizing the Entry of Transforming Deoxyribonucleic Acid into *Bacillus subtilis*. . *Journal of Bacteriology*, 89:281–287, February 1965. 12, 29, 99

[47] Norman Strauss. Early energy-dependent step in the entry of transforming deoxyribonucleic acid. *Journal of Bacteriology*, 101(1):35–37, January 1970. 12, 14, 29, 99

[48] Kenneth Briley Jr, Angella Dorsey Oresto, Peter Prepiak, Miguel Dias, Jessica Mann, and David Dubnau. The secretion ATPase ComGA is required for the binding and transport of transforming DNA. *Molecular Microbiology*, 81(3):818–830, 2011. 12, 29, 99, 100, 104

[49] J A Londoño-Vallejo and D Dubnau. Mutation of the putative nucleotide binding site of the Bacillus subtilis membrane protein ComFA abolishes the uptake of DNA during transformation. *Journal of Bacteriology*, 176(15):4642–4645, August 1994. 12, 14, 17, 21, 25, 26, 29, 47, 102, 109, 120, 121, 124

[50] Naomi Kramer, Jeanette Hahn, and David Dubnau. Multiple interactions among the competence proteins of Bacillus subtilis. *Molecular Microbiology*, 65(2):454–464, July 2007. 13, 70, 105, 138

[51] Willem M Vos, Gerard Venema, Umberto Canosi, and Thomas A Trautner. Plasmid transformation in Bacillus subtilis: Fate of plasmid DNA. *MGG Molecular & General Genetics*, 181(4):424–433, 1981. 13

[52] M Albano, R Breitling, and DA Dubnau. Nucleotide sequence and genetic organization of the Bacillus subtilis comG operon. *Journal of Bacteriology*, 171(10):5386–5404, 1989. 13

[53] Y S Chung and D Dubnau. All seven comG open reading frames are required for DNA binding during transformation of competent Bacillus subtilis. *Journal of Bacteriology*, 180(1):41–45, January 1998. 13, 100

[54] R Provvedi and D Dubnau. ComEA is a DNA receptor for transformation of competent Bacillus subtilis. *Molecular Microbiology*, 31(1):271–280, January 1999. 13, 14, 70

[55] G Inamine and D Dubnau. ComEA, a Bacillus subtilis integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *Journal of Bacteriology*, 177(11):3045–3051, June 1995. 14, 70, 100

[56] Masaomi Takeno, Hisataka Taguchi, and Takashi Akamatsu. Role of ComEA in DNA uptake during transformation of competent Bacillus subtilis. *JBIOSC*, pages 1–5, February 2012. 14

[57] Irena Draskovic and David Dubnau. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Molecular Microbiology*, 55(3):881–896, February 2005. 14, 105

[58] J A Londoño-Vallejo and D Dubnau. comF, a Bacillus subtilis late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Molecular Microbiology*, 9(1):119–131, July 1993. 14, 16, 17, 20, 21, 24, 26, 29, 47, 70, 104, 106, 111, 145

[59] J A Londoño-Vallejo and D Dubnau. Membrane association and role in DNA uptake of the Bacillus subtilis PriA analogue ComF1. *Molecular Microbiology*, 13(2):197–205, July 1994. 14, 53, 105, 109

[60] Masaomi Takeno, Hisataka Taguchi, and Takashi Akamatsu. Role of ComFA in controlling the DNA uptake rate during transformation of competent Bacillus subtilis. *Journal of bioscience and bioengineering*, March 2011. 14, 24

[61] R Davidoff-Abelson and D Dubnau. Conditions affecting the isolation from transformed cells of Bacillus subtilis of high-molecular-weight single-stranded deoxyribonucleic acid of donor origin. *Journal of Bacteriology*, 116(1):146–153, 1973. 17

[62] Narendra Tuteja and Renu Tuteja. Unraveling DNA helicases. Motif, structure, mechanism and function. *European journal of biochemistry / FEBS*, 271(10):1849–1863, May 2004. 17, 18, 19

[63] Mohamed Abdelhaleem. Helicases: an overview. *Methods in molecular biology (Clifton, NJ)*, 587:1–12, January 2010. 17, 18

[64] Eric Paget and Pascal Simonet. On the track of natural transformation in soil. *FEMS microbiology ecology*, 15(1):109–117, 1994. 17

[65] W M Toone, K E Rudd, and J D Friesen. deaD, a new Escherichia coli gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in rpsB, the gene encoding ribosomal protein S2. *Journal of Bacteriology*, 173(11):3291–3302, June 1991. 18, 23

[66] Margaret E Fairman-Williams, Ulf-Peter Guenther, and Eckhard Jankowsky. SF1 and SF2 helicases: family matters. *Current opinion in structural biology*, 20(3):313–324, June 2010. 18, 19, 25

[67] Timothy M Lohman, Eric J Tomko, and Colin G Wu. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nature reviews Molecular cell biology*, 9(5):391–401, May 2008. 18, 19

[68] Anna Marie Pyle. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual review of biophysics*, 37:317–336, January 2008. 18, 19

[69] S W Matson and J W George. DNA helicase II of Escherichia coli. Characterization of the single-stranded DNA-dependent NTPase and helicase activities. *The Journal of biological chemistry*, 262(5):2066–2076, February 1987. 18

[70] S R Schmid and P Linder. Translation initiation factor 4A from Saccharomyces cerevisiae: analysis of residues conserved in the D-E-A-D family of RNA helicases. *Molecular and Cellular Biology*, 11(7):3463–3471, July 1991. 18

[71] Eric J Enemark and Leemor Joshua-Tor. On helicases and other motor proteins. *Current opinion in structural biology*, 18(2):243–257, April 2008. 18

[72] Martin R Singleton and Dale B Wigley. Modularity and specialization in superfamily 1 and 2 helicases. *Journal of Bacteriology*, 184(7):1819–1826, April 2002. 19

[73] Arnon Henn, Wenxiang Cao, Nicholas Licciardello, Sara Heitkamp, David Hackney, and Enrique M De La Cruz. Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. *Proceedings of the National Academy of Sciences*, 107(9):4046–4050, 2010. 20, 27

[74] J Hahn, M Albano, and D Dubnau. Isolation and characterization of Tn917lac-generated competence mutants of Bacillus subtilis. *Journal of Bacteriology*, 169(7):3104–3109, July 1987. 20

[75] C Notredame, D G Higgins, and J Heringa. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302(1):205–217, September 2000. 23

[76] Iain M Wallace, Orla O'Sullivan, Desmond G Higgins, and Cedric Notredame. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic acids research*, 34(6):1692–1699, 2006. 23

[77] Sebastien Moretti, Fabrice Armougom, Iain M Wallace, Desmond G Higgins, Cornelius V Jongeneel, and Cedric Notredame. The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic acids research*, 35(Web Server issue):W645–8, July 2007. 23

[78] Paolo Di Tommaso, Sebastien Moretti, Ioannis Xenarios, Miquel Orobitg, Alberto Montanyola, Jia-Ming Chang, Jean-Francois Taly, and Cedric Notredame. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research*, 39(Web Server issue):W13–7, July 2011. 23, 50

[79] R Godbout and J Squire. Amplification of a DEAD box protein gene in retinoblastoma cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16):7578–7582, August 1993. 23

[80] P Linder and P P Slonimski. Sequence of the genes TIF1 and TIF2 from Saccharomyces cerevisiae coding for a translation initiation factor. *Nucleic acids research*, 16(21):10359, November 1988. 23

[81] K Nishi, F Morel-Deville, J W Hershey, T Leighton, and J Schnier. An eIF-4A-like protein is a suppressor of an Escherichia coli mutant defective in 50S ribosomal subunit assembly. *Nature*, 336(6198):496–498, December 1988. 23

[82] K Theis, P.J Chen, M Skorvaga, B Van Houten, and C Kisker. Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. *The EMBO journal*, 18(24):6899–6907, 1999. 23

[83] S B Inoue, H Sakamoto, H Sawa, and Y Shimura. Nucleotide sequence of a fission yeast gene encoding the DEAH-box RNA helicase. *Nucleic acids research*, 20(21):5841, November 1992. 23

[84] D Foulger and J Errington. A 28 kbp segment from the spoVM region of the Bacillus subtilis 168 genome. *Microbiology (Reading, England)*, 144 (Pt 3):801–805, March 1998. 23

[85] M D Adams, S E Celniker, R A Holt, C A Evans, J D Gocayne, P G Amanatides, S E Scherer, P W Li, R A Hoskins, R F Galle, R A George, S E Lewis, S Richards, M Ashburner, S N Henderson, G G Sutton, J R Wortman, M D Yandell, Q Zhang, L X Chen, R C Brandon, Y H Rogers, R G Blazej, M Champe, B D Pfeiffer, K H Wan, C Doyle, E G Baxter, G Helt, C R Nelson, G L Gabor, J F Abril, A Agbayani, H J An, C Andrews-Pfannkoch, D Baldwin, R M Ballew, A Basu, J Baxendale, L Bayraktaroglu, E M Beasley, K Y Beeson, P. V Benos, B P Berman, D Bhandari, S Bolshakov, D Borkova, M R Botchan, J Bouck, P Brokstein, P Brottier, K C Burtis, D A Busam, H Butler, E Cadieu, Center, A, I Chandra, J M Cherry, S Cawley, C Dahlke, L B Davenport, P Davies, B de Pablos, A Delcher, Z Deng, A D Mays, I Dew, S M Dietz, K Dodson, L E Doup, M Downes, S Dugan-Rocha, B C Dunkov, P Dunn, K J Durbin, C C Evangelista, C Ferraz, S Ferriera, W Fleischmann, C Fosler, A E Gabrielian, N S Garg, W M Gelbart, K Glasser, A Glodek, F Gong, J H Gorrell, Z Gu, P Guan, M Harris, N L Harris, D Harvey, T J Heiman, J R Hernandez, J Houck, D Hostin, K A Houston, T J Howland, M H Wei, C Ibegwam, M Jalali, F Kalush, G H Karpen, Z Ke, J A Kennison, K A Ketchum, B E Kimmel, C D Kodira, C Kraft, S Kravitz, D Kulp, Z Lai, P Lasko, Y Lei, A A Levitsky, J Li, Z Li, Y Liang, X Lin, X Liu, B Mattei, T C McIntosh, M P McLeod, D McPherson, G Merkulov, N V Milshina, C Mobarry, J Morris, A Moshrefi, S M Mount, M Moy, B Murphy, L Murphy, D M Muzny, D L Nelson, D R Nelson, K A Nelson, K Nixon, D R Nusskern, J M Pacleb, M Palazzolo, G S Pittman, S Pan, J Pollard, V Puri, M G Reese, K Reinert, K Remington, R D Saunders, F Scheeler, H Shen, B C Shue, I Siden-Kiamos, M Simpson, M P Skupski, T Smith, E Spier, A C Spradling, M Stapleton, R Strong, E Sun, R Svirskas, C Tector, R Turner, E Venter, A H Wang, X Wang, Z Y Wang, D A Wassarman, G M Weinstock, J Weissenbach, S M Williams, WoodageT, K C Worley, D Wu, S Yang, Q A Yao, J Ye, R F Yeh, J S Zaveri, M Zhan, G Zhang, Q Zhao, L Zheng, X H Zheng, F N Zhong, W Zhong, X Zhou, S Zhu, X Zhu, H O Smith, R A Gibbs, E W Myers, G M Rubin, and J C Venter. The genome sequence of Drosophila melanogaster. *Science*, 287(5461):2185–2195, March 2000. 23

[86] Josette Banroques, Monique Doere, Marc Dreyfus, Patrick Linder, and N Kyle Tanner. Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *Journal of molecular biology*, 396(4):949–966, March 2010. 25

[87] Zhizhong Gong, Chun-Hai Dong, Hojoung Lee, Jianhua Zhu, Liming Xiong, Deming Gong, Becky Stevenson, and Jian-Kang Zhu. A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in Arabidopsis. *The Plant Cell Online*, 17(1):256–267, 2005. 25

[88] P Linder. Dead-box proteins: a family affair–active and passive players in RNP-remodeling. *Nucleic acids research*, January 2006. 25

[89] N Kyle Tanner, Olivier Cordin, Josette Banroques, Monique Doere, and Patrick Linder. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Molecular cell*, 11(1):127–138, January 2003. 25, 120

[90] Norman Strauss. Configuration of Transforming Deoxyribonucleic Acid During Entry into *Bacillus subtilis*. *Journal of Bacteriology*, 89:288–293, February 1965. 28

[91] Norman Strauss. Further evidence concerning the configuration of transforming deoxyribonucleic acid during entry into Bacillus subtilis. *Journal of Bacteriology*, 91(2):702–708, February 1966. 28

[92] P J Youngman, J B Perkins, and R Losick. Genetic transposition and insertional mutagenesis in Bacillus subtilis with Streptococcus faecalis transposon Tn917. *Proceedings of the National Academy of Sciences of the United States of America*, 80(8):2305–2309, April 1983. 30, 39, 72, 87, 124, 133, 146

[93] M Konkol, K Blair, and D Kearns. Plasmid-Encoded ComI Inhibits Competence in the Ancestral 3610 Strain of Bacillus subtilis. *Journal of Bacteriology*, January 2013. 30, 72, 124, 146

[94] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6(5):343–345, May 2009. 30, 122

[95] Katherine P Lemon and Alan D Grossman. Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science*, 282(5393):1516–1519, 1998. 30, 40, 150

[96] Jeff Yon and Mike Fried. Precise gene fusion by PCR. *Nucleic acids research*, 17(12):4895, 1989. 31

[97] J Patrick and D Kearns. MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. *Molecular Microbiology*, January 2008. 35, 40, 88, 122, 134

[98] David Giedroc, Kathleen Keating, Craig Martin, Kenneth Williams, and Joseph Coleman. Zinc metalloproteins involved in replication and transcription. *Journal of inorganic biochemistry*, 28(2):155–169, 1986. 45

[99] J M Berg and Y Shi. The Galvanization of Biology: A Growing Appreciation for the Roles of Zinc. *Science*, 271(5252):1081–1085, February 1996. 45

[100] S Durai, M Mani, K Kandavelou, J Wu, M.H Porteus, and S Chandrasegaran. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic acids research*, 33(18):5978–5990, 2005. 45

[101] J.L Liu, P Rigolet, S.X Dou, P.Y Wang, and X G Xi. The zinc finger motif of Escherichia coli RecQ is implicated in both DNA binding and protein folding. *Journal of Biological Chemistry*, 279(41):42794, 2004. 45, 46, 58, 86

[102] S Esposito, I Baglivo, G Malgieri, L Russo, L Zaccaro, L.D D'Andrea, M Mammucari, B Di Blasio, C Isernia, and R Fattorusso. A novel type of zinc finger DNA binding domain in the Agrobacterium tumefaciens transcriptional regulator Ros. *Biochemistry*, 45(34):10394–10405, 2006. 45

[103] W S El-Deiry, K M Downey, and A G So. Molecular mechanisms of manganese mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 81(23):7378–7382, December 1984. 45

[104] C B Black, H W Huang, and J A Cowan. Biological coordination chemistry of magnesium, sodium, and potassium ions. Protein and nucleotide binding sites. *Coordination Chemistry Reviews*, 1994. 46

[105] LubomÄ±Ìr RulÄ±Ìšek and JiřÄ±Ì Vondrášek. Coordination geometries of selected transition metal ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+) in metalloproteins. *Journal of inorganic biochemistry*, 71(3-4):115–127, September 1998. 46

[106] R Choudhury and S Srivastava. Zinc resistance mechanisms in bacteria. *Current Science*, 81(7):768–775, 2001. 46

[107] P.V Pedone, R Ghirlando, G.M Clore, A.M Gronenborn, G Felsenfeld, and J.G Omichinski. The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. *Proceedings of the National Academy of Sciences*, 93(7):2822–2826, 1996. 46, 70

[108] Y Liu, L Yu, W Yu, S Shi, B Sun, G Wu, and Z Huang. Expression and purification of a novel ZNF191 zinc finger protein. *Science in China Series B: Chemistry*, 42(3):245–252, 1999. 46

[109] Elmar Nurmemmedov and Marjolein Thunnissen. Expression, purification, and characterization of the 4 zinc finger region of human tumor suppressor WT1. *Protein expression and purification*, 46(2):379–389, April 2006. 46

[110] Lei Jiang, Daolin Tang, Kangkai Wang, Huali Zhang, Can Yuan, Dayue Duan, and Xianzhong Xiao. Functional analysis of a novel KRAB/C2H2 zinc finger protein Mipu1. *Biochemical and biophysical research communications*, 356(4):829–835, May 2007. 46

[111] Wolfgang Maret. New perspectives of zinc coordination environments in proteins. *Journal of inorganic biochemistry*, 111:110–116, 2012. 46

[112] A. V Persikov and M Singh. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic acids research*, 42(1):97–108, January 2014. 46

[113] Marco Pagni, Vassilios Ioannidis, Lorenzo Cerutti, Monique Zahn-Zabal, C Victor Jongeneel, and Laurent Falquet. MyHits: a new interactive resource for protein annotation and domain identification. *Nucleic acids research*, 32(Web Server issue):W332–5, July 2004. 46, 48, 106, 114

[114] J Cheng, A Z Randall, M J Sweredoski, and P Baldi. SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research*, 33(Web Server issue):W72–6, July 2005. 46, 48, 58

[115] Marco Pagni, Vassilios Ioannidis, Lorenzo Cerutti, Monique Zahn-Zabal, C Victor Jongeneel, Jörg Hau, Olivier Martin, Dmitri Kuznetsov, and Laurent Falquet. MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic acids research*, 35(Web Server issue):W433–7, July 2007. 46, 48, 106, 114

[116] N Shu, T Zhou, and S Hovmöller. Prediction of zinc-binding sites in proteins from sequence. *Bioinformatics*, 24(6):775–782, 2008. 46, 48

[117] Zhen Chen, Yanying Wang, Ya-Feng Zhai, Jiangning Song, and Ziding Zhang. ZincExplorer: an accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Molecular bioSystems*, 9(9):2213–2222, 2013. 46, 48

[118] Carsten Kemena and Cedric Notredame. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics*, 25(19):2455–2465, October 2009. 50

[119] H Takami, K Nakasone, Y Takaki, G Maeno, R Sasaki, N Masui, F Fuji, C Hirama, Y Nakamura, N Ogasawara, S Kuhara, and K Horikoshi. Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis. *Nucleic acids research*, 28(21):4317–4331, November 2000. 50

[120] A Bolotin, P Wincker, S Mauger, O Jaillon, K Malarme, J Weissenbach, S D Ehrlich, and A Sorokin. The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. *Genome research*, 11(5):731–753, May 2001. 50

[121] J J Ferretti, W M McShan, D Ajdic, D J Savic, G Savic, K Lyon, C Primeaux, S Sezate, A N Suvorov, S Kenton, H S Lai, S P Lin, Y Qian, H G Jia, F Z Najar, Q Ren, H Zhu, L Song, J White, X Yuan, S W Clifton, B A Roe, and R McLaughlin. Complete genome sequence of an M1 strain of Streptococcus pyogenes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4658–4663, April 2001. 50

[122] H Tettelin, K E Nelson, I T Paulsen, J A Eisen, T D Read, S Peterson, J Heidelberg, R T DeBoy, D H Haft, R J Dodson, A S Durkin, M Gwinn, J F Kolonay, W C Nelson, J D Peterson, L A Umayam, O White, S L Salzberg, M R Lewis, D Radune, E Holtzapple, H Khouri, A M Wolf, T R Utterback, C L Hansen, L A McDonald, T V Feldblyum, S Angiuoli, T Dickinson, E K Hickey, I E Holt, B J Loftus, F Yang, H O Smith, J C Venter, B A Dougherty, D A Morrison, S K Hollingshead, and C M Fraser. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science*, 293(5529):498–506, July 2001. 50

[123] P Glaser, L Frangeul, C Buchrieser, C Rusniok, A Amend, F Baquero, P Berche, H Bloecker, P Brandt, T Chakraborty, A Charbit, F Chetouani, E Couve, A de Daruvar, P Dehoux, E Domann, G Dominguez-Bernal, E Duchaud, L Durant, O Dussurget, K D Entian, H Fsihi, F Garcia-del Portillo, P Garrido, L Gautier, W Goebel, N Gomez-Lopez, T Hain, J Hauf, D Jackson, L M Jones, U Kaerst, J Kreft, M Kuhn, F Kunst, G Kurapkat, E Madueno, A Maitournam, J M Vicente, E Ng, H Nedjari, G Nordsiek, S Novella, B de Pablos, J C Perez-Diaz, R Purcell, B Remmel, M Rose, T Schlueter, N Simoes, A Tierrez, J A Vazquez-Boland, H Voss, J Wehland, and P Cossart. Comparative genomics of Listeria species. *Science*, 294(5543):849–852, October 2001. 50

[124] Laura Garcia-Alvarez, Matthew T G Holden, Heather Lindsay, Cerian R Webb, Derek F J Brown, Martin D Curran, Enid Walpole, Karen Brooks, Derek J Pickard, Christopher Teale, Julian Parkhill, Stephen D Bentley, Giles F Edwards, E Kirsty Girvan, Angela M Kearns, Bruno Pichon, Robert L R Hill, Anders Rhod Larsen, Robert L Skov, Sharon J Peacock, Duncan J Maskell, and Mark A Holmes. Meticillin-resistant Staphylococcus aureus with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. *The Lancet infectious diseases*, 11(8):595–603, August 2011. 50

[125] Richard Burgess. Protein precipitation techniques. *Methods in enzymology*, 463:331–342, 2009. 53

[126] Richard Burgess. Refolding solubilized inclusion body proteins. *Methods Enzymol*, 463(11):259–282, 2009. 53

[127] L E Leong. The use of recombinant fusion proteases in the affinity purification of recombinant proteins. *Molecular biotechnology*, 12(3):269–274, October 1999. 62, 64

[128] David Auld. Zinc coordination sphere in biochemical zinc sites. *Biometals*, 14(3-4):271–313, 2001. 67, 70

[129] A Soltyk, D Shugar, and M Piechowska. Heterologous deoxyribonucleic acid uptake and complexing with cellular constituents in competent Bacillus subtilis. *Journal of Bacteriology*, 124(3):1429–1438, December 1975. 70, 94

[130] J Hahn, B Maier, BJ Haijema, M Sheetz, and D Dubnau. Transformation proteins and DNA uptake localize to the cell poles in Bacillus subtilis. *Cell*, 122(1):59–71, 2005. 70, 105, 138

[131] M Kaufenstein, M van der Laan, and P.L Graumann. The three-layered DNA uptake machinery at the cell pole in competent Bacillus subtilis cells is a stable complex. *Journal of Bacteriology*, 193(7):1633–1642, 2011. 70, 105, 138

[132] Richard Simpson. Precipitation of Proteins by Polyethylenimine. *Cold Spring Harbor Protocols*, 2006(1):pdb.prot4312, June 2006. 84

[133] Irena Vorá̌čková, Sárka Suchanová, Pavel Ulbrich, William E Diehl, and Tomáš Ruml. Purification of proteins containing zinc finger domains using immobilized metal ion affinity chromatography. *Protein expression and purification*, May 2011. 86

[134] S Vilain, J M Pretorius, J Theron, and V S Brozel. DNA as an Adhesin: Bacillus cereus Requires Extracellular DNA To Form Biofilms. *Applied and Environmental Microbiology*, 75(9):2861–2868, April 2009. 93

[135] Olga Zafra, María Lamprecht-Grandío, Carolina González de Figueras, and José Eduardo González-Pastor. Extracellular DNA release by undomesticated Bacillus subtilis is regulated by early competence. *PLoS ONE*, 7(11):e48716, 2012. 93

[136] Bea Baur, Kurt Hanselmann, Wolfram Schlimme, and Bernard Jenni. Genetic transformation in freshwater: Escherichia coli is able to develop natural competence. *Applied and Environmental Microbiology*, 62(10):3673–3678, 1996. 93

[137] J Hahn, A Luttinger, and D Dubnau. Regulatory inputs for the synthesis of ComK, the competence transcription factor of Bacillus subtilis. *Molecular Microbiology*, 21(4):763–775, August 1996. 93

[138] D A Morrison and M S Lee. Regulation of competence for genetic transformation in Streptococcus pneumoniae: a link between quorum sensing and DNA processing genes. *Research in Microbiology*, 151(6):445–451, January 2000. 93

[139] L P MacFadyen, D Chen, H C Vo, D Liao, R Sinotte, and R J Redfield. Competence development by Haemophilus influenzae is regulated by the availability of nucleic acid precursors. *Molecular Microbiology*, 40(3):700–707, May 2001. 93

[140] X Charpentier, E Kay, D Schneider, and H. A Shuman. Antibiotics and UV Radiation Induce Competence for Natural Transformation in Legionella pneumophila. *Journal of Bacteriology*, 193(5):1114–1121, March 2011. 93

[141] S Yamamoto, H Izumiya, J Mitobe, M Morita, E Arakawa, M Ohnishi, and H Watanabe. Identification of a Chitin-Induced Small RNA That Regulates Translation of the tfoX Gene, Encoding a Positive Regulator of Natural Competence in Vibrio cholerae. *Journal of Bacteriology*, 193(8):1953–1965, April 2011. 93

[142] M Piechowska and M S Fox. Fate of transforming deoxyribonucleate in Bacillus subtilis. *Journal of Bacteriology*, 108(2):680–689, November 1971. 94

[143] David Dubnau and Carol Cirigliano. Fate of transforming DNA following uptake by competent Bacillus subtilis: III. Formation and properties of products isolated from transformed cells which are derived entirely from donor DNA. *Journal of molecular biology*, 64(1):9–29, 1972. 94

[144] R Davidoff-Abelson and D Dubnau. Kinetic analysis of the products of donor deoxyribonucleate in transformed cells of Bacillus subtilis. *Journal of Bacteriology*, 116(1):154–162, October 1973. 94

[145] H Joenje and G Venema. Different nuclease activities in competent and noncompetent Bacillus subtilis. *Journal of Bacteriology*, 122(1):25–33, April 1975. 94

[146] Candela Manfredi, Begoña Carrasco, Silvia Ayora, and Juan C Alonso. Bacillus subtilis RecO nucleates RecA onto SsbA-coated single-stranded DNA. *The Journal of biological chemistry*, 283(36):24837–24847, September 2008. 94

[147] G Pozzi, L Masala, F Iannelli, R Manganelli, L S Havarstein, L Piccoli, D Simon, and D A Morrison. Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone. *Journal of Bacteriology*, 178(20):6087–6090, October 1996. 95

[148] Finn Aas, Matthew Wolfgang, Stephan Frye, Steven Dunham, Cecilia Løvold, and Michael Koomey. Competence for natural transformation in Neisseria gonorrhoeae: components of DNA binding and uptake linked to type IV pilus expression. *Molecular Microbiology*, 46(3):749–760, 2002. 95

[149] K Stingl, S Muller, G Scheidgen-Kleyboldt, M Clausen, and B Maier. Composite system mediates two-step DNA uptake into Helicobacter pylori. *Proceedings of the National Academy of Sciences*, 107(3):1184–1189, January 2010. 99, 100

[150] Patricia Melin, Vincent Thoreau, Caroline Norez, Frédéric Bilan, Alain Kitzis, and Frédéric Becq. The cystic fibrosis mutation G1349D within the signature motif LSHGH of NBD2 abolishes the activation of CFTR chloride channels by genistein. *Biochemical pharmacology*, 67(12):2187–2196, 2004. 118

[151] Antresh Kumar, Suneet Shukla, Ajeet Mandal, Sudhanshu Shukla, Suresh Ambudkar, and Rajendra Prasad. Divergent signature motifs of nucleotide binding domains of ABC multidrug transporter, CaCdr1p of pathogenic Candida albicans, are functionally asymmetric and noninterchangeable. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1798(9):1757–1766, 2010. 118

[152] K. A Bernstein, S Granneman, A. V Lee, S Manickam, and S. J Baserga. Comprehensive Mutational Analysis of Yeast DEXD/H Box RNA Helicases Involved in Large Ribosomal Subunit Biogenesis. *Molecular and Cellular Biology*, 26(4):1195–1208, February 2006. 120

[153] N Kyle Tanner. The newly identified Q motif of DEAD box helicases is involved in adenine recognition. *Cell Cycle*, 2(1):18–19, January 2003. 120

[154] Matthew T Cabeen, Godefroid Charbon, Waldemar Vollmer, Petra Born, Nora Ausmees, Douglas B Weibel, and Christine Jacobs-Wagner. Bacterial cell curvature through mechanical control of cell growth. *The EMBO journal*, 28(9):1208–1219, May 2009. 142